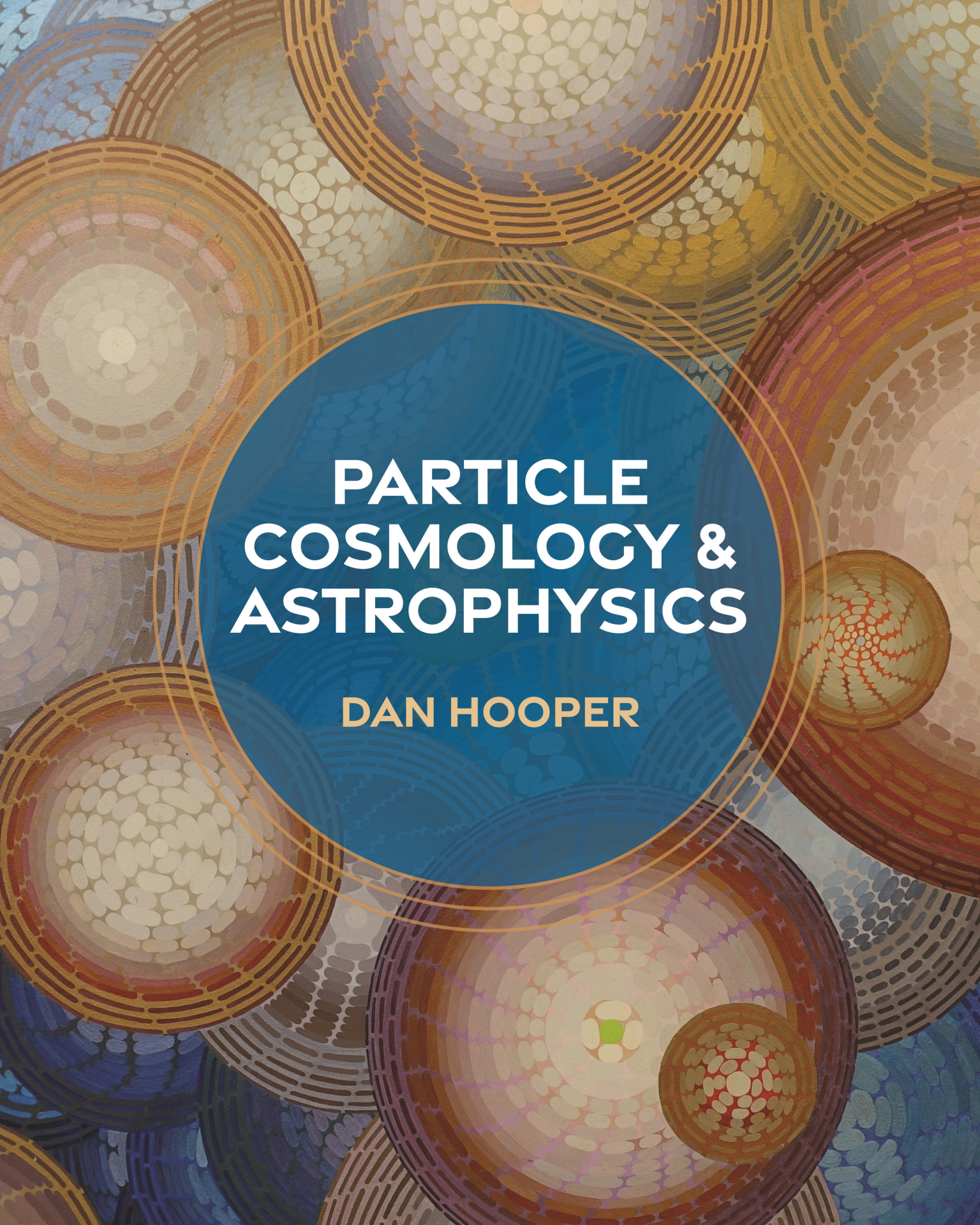

PARTICLE
COSMOLOGY &
ASTROPHYSICS
DAN HOOPER

Particle Cosmology and Astrophysics

Particle Cosmology and Astrophysics

Dan Hooper

PRINCETON UNIVERSITY PRESS

PRINCETON AND OXFORD

Copyright © 2024 by Princeton University Press

Princeton University Press is committed to the protection of copyright and the intellectual property our authors entrust to us. Copyright promotes the progress and integrity of knowledge. Thank you for supporting free speech and the global exchange of ideas by purchasing an authorized edition of this book. If you wish to reproduce or distribute any part of it in any form, please obtain permission.

Requests for permission to reproduce material from this work
should be sent to permissions@press.princeton.edu

Published by Princeton University Press
41 William Street, Princeton, New Jersey 08540
99 Banbury Road, Oxford OX2 6JX

press.princeton.edu

All Rights Reserved
ISBN 978-0-691-23504-2
ISBN (e-book) 978-0-691-23506-6

British Library Cataloging-in-Publication Data is available

Editorial: Ingrid Gnerlich and Whitney Rauenhorst
Production Editorial: Jenny Wolkowicki
Cover design: Wanda España
Production: Jacqueline Poirier
Copyeditor: Bhisham Bherwani

Cover image: Angela Lahs-Gonzales, detail from *Untitled*, c. 1970s, oil on masonite, 42 × 42 inches.
Image courtesy of the Angela Lahs-Gonzales Archive.

This book has been composed in Minionpro

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To Rocky and Mike, for blazing one hell of a trail.

Contents

Foreword xiii
Edward W. Kolb and Michael Turner

Constants, Parameters, and Reference Tables xv

Frequently Used Equations xxiii

Preface xxv

I Preliminaries 1

1 Our Universe 3
1.1 A brief history of our universe 4
1.2 The empirical pillars of modern cosmology 6
1.3 Cosmic concordance 8
1.4 Open questions and outstanding puzzles 8

2 General Relativity and the Friedmann Equations 11
2.1 Spacetime and the metric 11
2.2 Tensor notation in general relativity 14
2.3 The friedmann equations 15
2.4 Matter, radiation, and pressure 18
2.5 Static universes and the cosmological constant 19
2.6 The evolution of energy 21
2.7 The expansion history of the universe 22
2.8 Distances in an expanding universe 25
Problems 28

3 A Crash Course in Particle Physics 30
3.1 The Standard Model of particle physics 30
3.2 Feynman diagrams 31

3.3 Conservation laws 35
3.4 Phase transitions in the Standard Model 37
3.5 Amplitudes 39
3.6 Relativistic kinematics and four-vectors 40
3.7 Decays and scattering 42
Problems 44

4 Thermodynamics for Cosmologists 46
4.1 Equilibrium thermodynamics 46
4.2 Entropy 50
4.3 Chemical potentials 51
4.4 Departures from equilibrium 53
4.5 Neutrino decoupling 54
4.6 Evolution after decoupling 56
4.7 The boltzmann equation 57
Problems 63

5 The Interactions of High-Energy Particles in Astrophysics 65
5.1 Interactions of high-energy electrons 65
5.2 Interactions of high-energy nucleons and nuclei 71
5.3 Interactions of high-energy photons 74
Problems 76

II Cosmology 77

6 Recombination and Photon Decoupling 79
6.1 Photon scattering 79
6.2 The ionization history of the early universe 81
6.3 Reionization 82
6.4 The surface of last scattering 84
Problems 85

7 The Cosmic Microwave Background 86
7.1 The spectrum of the CMB 86
7.2 Temperature anisotropies 88
7.3 The angular power spectrum of the CMB 88
7.4 The evolution of density perturbations 92
7.5 Features of the CMB's angular power spectrum 94
7.6 Extracting cosmological parameters 98
7.7 CMB polarization 101
7.8 The current state of CMB measurements 102

7.9 Dark radiation 104
Problems 106

8 Big Bang Nucleosynthesis 108
8.1 Proton and neutron freeze out 109
8.2 Nuclear abundances in equilibrium 112
8.3 From nucleons to helium 114
8.4 Beyond helium 116
8.5 The primordial element abundances 118
8.6 BBN as a probe of exotic physics 119
Problems 122

9 Dark Matter 124
9.1 Dynamical evidence for dark matter 124
9.2 MACHOs and gravitational microlensing 125
9.3 What about modified gravity? 127
9.4 Requirements of a dark matter candidate 127
9.5 The abundance of a thermal relic 129
9.6 Asymmetric dark matter 139
9.7 Nonthermal production mechanisms 141
Problems 147

10 The Formation of Large-Scale Structure 150
10.1 Gravitational collapse 150
10.2 Pressure and the Jeans length 152
10.3 The evolution of structure in an expanding universe 155
10.4 The matter power spectrum 157
10.5 From density perturbations to dark matter halos 159
10.6 Hot, warm, and cold dark matter 164
10.7 Baryon acoustic oscillations 168
10.8 Adiabatic and isocurvature perturbations 170
Problems 171

11 Neutrino Masses and Oscillations 173
11.1 Neutrino masses 173
11.2 Neutrino mixing and oscillations 176
11.3 Neutrino oscillations in the presence of matter 179
Problems 181

12 Baryogenesis and Leptogenesis 182
12.1 The matter-antimatter asymmetry 182
12.2 The Sakharov conditions 184

12.3 Baryogenesis in grand unified theories 185
12.4 Sphalerons 190
12.5 Baryogenesis via leptogenesis 194
12.6 Electroweak baryogenesis 197
12.7 Topological defects 202
Problems 207

13 Inflation 210
13.1 Motivations for inflation 210
13.2 A common solution: accelerated expansion 213
13.3 The dynamics of a slowly rolling scalar field 215
13.4 Primordial perturbations from inflation 218
13.5 Gravitational waves and primordial *B*-modes 220
13.6 Examples of inflationary models 220
13.7 Reheating 224
13.8 Present and future probes of inflation 226
13.9 Eternal inflation 226
Problems 228

14 Dark Energy 230
14.1 The cosmological constant problem 230
14.2 Dark energy and accelerating expansion 231
14.3 Dynamical dark energy 231
14.4 Observational probes of dark energy 234
14.5 Dark energy and the anthropic principle 236
Problems 237

III Particle Astrophysics 239

15 Cosmic Rays 241
15.1 The cosmic-ray spectrum 241
15.2 Cosmic-ray propagation 242
15.3 Cosmic-ray acceleration 245
15.4 Ultra-high energy cosmic rays 248
Problems 251

16 Gamma-Ray and Neutrino Astrophysics 252
16.1 Why gamma rays? Why neutrinos? 252
16.2 The astrophysical production and propagation of gamma rays 253
16.3 The astrophysical production and propagation of neutrinos 257

16.4 Astrophysical sources of high-energy gamma rays and neutrinos 258
Problems 269

17 Particle Dark Matter Candidates 271
17.1 Supersymmetric dark matter 271
17.2 Sterile neutrinos 281
17.3 Axions 288
17.4 Dark matter in hidden sectors 300
Problems 301

18 Direct Dark Matter Searches 302
18.1 Dark matter scattering with nuclei 302
18.2 Dark matter scattering cross sections 304
18.3 Direct detection constraints and their impact 308
18.4 Axion detection 311
Problems 312

19 Searches for Dark Matter Annihilation and Decay 314
19.1 Motivations for indirect searches 314
19.2 Dark matter annihilation and the CMB 315
19.3 Gamma rays from dark matter annihilation 316
19.4 Cosmic rays from dark matter annihilation 320
19.5 Decaying dark matter 323
Problems 324

20 Collider Physics for Astronomers 326
20.1 Why collide particles? 326
20.2 Electron and proton colliders 327
20.3 Particle detection at the LHC 329
20.4 Cross sections and event rates at the LHC 330
20.5 The LHC and the early universe 332
Problems 333

Bibliography 335

Solutions to Selected Problems 361

Index 409

Foreword

Edward W. Kolb and Michael Turner

When we wrote *The Early Universe* in 1989, particle cosmology was just emerging as a field. This was not many years after Allan Sandage described cosmology as the search for two numbers, H_0 and q_0. Steve Weinberg's text, *Gravitation and Cosmology*, published in 1972, was widely read by physicists—especially particle physicists—who wanted to learn about general relativity, and cosmology was one of the applications of general relativity he developed in some detail. In his book, Weinberg coined the phrase "Standard Model" for the Hot Big Bang and described it in detail back to 10^{-5} seconds after the bang. However, his description stopped there because cosmology "hit the wall:" the hadron wall. At the time, it was thought that the finite size of hadrons would cause them to overlap, and the rising spectrum of hadronic states would make a thermal description problematic.

All that changed with the discovery of quarks and asymptotic freedom, the emergence of the Standard Model of particle physics, and the heady speculations about grand unifications that quickly followed. Instead of an impenetrable hadron wall there was quark soup, and the early Universe was open for exploration! On the astronomy side, instrumentation innovations like charged-coupled devices and a proliferation of terrestrial and space telescopes across the electromagnetic spectrum were opening astronomy's early universe, redshifts beyond unity, and much more. This was the backdrop for *The Early Universe*.

Our graduate-level text described the burst of theoretical effort that led to the ideas that underpin ΛCDM, today's Standard Model of cosmology, and paved the way for the study of baryogenesis, phase transitions, particle dark matter, inflation, and even dark energy. The discovery of the anisotropy of the CMB in 1992 spurred the preface of our second edition, which foreshadowed the era of precision cosmology today. All of this expanded the vocabulary of cosmology from the descriptive (H_0, q_0, Ω_0, and t_0) to the big ideas of dark matter, dark energy, inflation, baryogenesis, and phase transitions, while linking the inner space of elementary particles and the outer space of astrophysics and cosmology.

Less than 10 years later, the acceleration of the Universe had been discovered, the ΛCDM model was in place, and a deluge of cosmological data—most especially the CMB—was confirming the model and providing precision determinations of cosmological parameters. Citing just one example, the 50-sigma discrepancy between the baryon mass density and that of dark matter anchors the case for nonbaryonic dark matter.

Today, cosmology is done on an industrial scale by more than a thousand cosmologists, astronomers, and physicists. The level of activity, theoretical, observational, and experimental, is much more than we could have ever imagined: speculations about the

origin of space and time and the deep connections between gravity, quantum mechanics, and field theory; many new ideas about dark matter and dark sectors; the discovery of the Higgs boson and the search for additional new particles at accelerators; a broad exploration of neutrino masses, mixing angles, and CP violation; dark matter searches for WIMPs, axions, and a plethora of new candidates; galaxy surveys, CMB experiments with nanoKelvin sensitivity, and ground and space-based telescopes exploring the nearby Universe and to redshifts of 10 and beyond. And there is much, much more planned for the future.

A new textbook is needed for those just entering and those already working in the field of particle cosmology and astrophysics, with its diversity of early-career and senior scientists, astronomers and particle physicists, relativists and AMO physicists, and nuclear physicists and cosmic-ray physicists. Interdisciplinary research is exciting, but it is also challenging. How does a collider theorist learn early-universe cosmology? General relativity is not usually part of the background of gamma-ray astrophysicists. And so on. Hooper's book covers all the basics needed to dive into this multidisciplinary research with stunning opportunities for discovery ahead. It will be an indispensable reference not only for those entering the field, but also for those who are already deeply involved. If you already can read a Feynman diagram, but not a diagram of an AGN jet, or vice versa, this is the book for you.

Beyond its use as a reference text, *Particle Cosmology and Astrophysics* will be an excellent textbook for advanced undergraduates or beginning graduate students. The material is presented at an accessible level, and the problems at the ends of the chapters challenge the readers' understanding of the material.

Rocky Kolb and Michael Turner

Constants, Parameters, and Reference Tables

Units

Throughout most of this book, I will adopt the system of natural units, in which $\hbar = c = k_B = 1$. Furthermore, I'll usually omit these symbols from the equations, except in those cases in which including them helps to make the meaning more clear. In the system of natural units, quantities of energy, mass, temperature, length^{-1}, and time^{-1} each carry the same units (that of energy). Some useful constants and conversion factors include:

Speed of light	$c = 2.998 \times 10^{10}$ cm/s
Reduced Planck's constant	$\hbar \equiv \dfrac{h}{2\pi} = 6.62607 \times 10^{-34}$ J s
	$= 6.582 \times 10^{-25}$ GeV s
Conversion factors	$\hbar c = 1.973 \times 10^{-14}$ GeV cm
	$1\,\text{J} = 6.242 \times 10^{9}$ GeV
	$1\,\text{kg} = 5.610 \times 10^{26}$ GeV
	$1\,\text{K} = 8.617 \times 10^{-14}$ GeV
	$1\,\text{parsec (pc)} = 3.086 \times 10^{18}$ cm

Cosmological parameters

Rate of Hubble expansion	$H_0 \approx 67 - 73$ km/s/Mpc
	$\approx (2.2 - 2.4) \times 10^{-18}$ s^{-1}
	$\approx (1.43 - 1.56) \times 10^{-42}$ GeV
Rate of Hubble expansion (alternative normalization)	$h \equiv \dfrac{H_0}{100\,\text{km/s/Mpc}} \approx 0.67 - 0.73$
Critical density	$\rho_{\text{crit}} \equiv \dfrac{3H_0^2 M_{\text{Pl}}^2}{8\pi} = 8.557 \times 10^{-33}$ kg/cm^3
	$= 4.944 \times 10^{-6}$ GeV/cm^3
	$= 1.302 \times 10^{11}\, M_\odot$/Mpc^3
	$= 3.797 \times 10^{-47}$ GeV^4
Dark matter density	$\Omega_{\text{DM}} \equiv \dfrac{\rho_{\text{DM}}}{\rho_{\text{crit}}} = 0.265$
Baryonic matter density	$\Omega_b \equiv \dfrac{\rho_b}{\rho_{\text{crit}}} = 0.049$
Total matter density	$\Omega_M = \Omega_b + \Omega_{\text{DM}} \equiv \dfrac{\rho_M}{\rho_{\text{crit}}} = 0.315$
Dark energy density	$\Omega_\Lambda \equiv \dfrac{\rho_\Lambda}{\rho_{\text{crit}}} = 0.685$
Photon (CMB) energy density	$\Omega_\gamma \equiv \dfrac{\rho_\gamma}{\rho_{\text{crit}}} = 5.382 \times 10^{-5}$
Neutrino energy density	$\Omega_\nu \equiv \dfrac{\rho_\nu}{\rho_{\text{crit}}} = 3.88 \times 10^{-5}$
	(for three massless neutrinos)
Total radiation density	$\Omega_{\text{rad}} = \Omega_\gamma + \Omega_\nu \equiv \dfrac{\rho_{\text{rad}}}{\rho_{\text{crit}}} = 9.26 \times 10^{-5}$
Baryon-to-photon ratio	$\eta \equiv \dfrac{n_b}{n_\gamma} \approx 6 \times 10^{-10}$
CMB temperature	$T_{\text{CMB}} = 2.7255$ K
	$= 0.000235$ eV
CMB number density	$n_{\text{CMB}} = 410.7$ cm^{-3}
Primordial helium mass fraction	$Y_p \equiv \dfrac{4n_{\text{He}}}{n_{\text{H}} + 4n_{\text{He}}} = 0.245$
Redshift at matter-radiation equality	$z_{\text{eq}} = 3402$
Time of matter-radiation equality	$t_{\text{eq}} = 5.11 \times 10^4$ yr

Age of the universe	$t_{\rm age} = 1.380 \times 10^{10}$ yr
Spectral index of primordial density fluctuations	$n_s = 0.965$
Amplitude of the linear matter power spectrum	$\sigma_8 = 0.81$
Reionization optical depth	$\tau = 0.054$

Astrophysical constants

Solar mass	$M_\odot = 1.989 \times 10^{30}\,\mathrm{kg}$
Solar luminosity	$L_\odot = 3.845 \times 10^{26}\,\mathrm{J/s}$
Solar radius	$R_\odot = 6.957 \times 10^{10}\,\mathrm{cm}$
Earth mass	$M_\oplus = 5.972 \times 10^{24}\,\mathrm{kg}$
Earth radius	$R_\oplus = 6.378 \times 10^{8}\,\mathrm{cm}$
Earth-sun distance (astronomical unit)	$1\,\mathrm{AU} = 1.496 \times 10^{13}\,\mathrm{cm}$
Solar distance from Galactic Center	$R_0 \approx 8250\,\mathrm{pc}$
Local dark matter density	$\rho_0 \approx 0.3 - 0.4\,\mathrm{GeV/cm^3}$

Physical constants

Planck mass	$M_{\rm Pl} \equiv \left(\frac{\hbar c}{G}\right)^{1/2} = 2.176 \times 10^{-8}\,{\rm kg}$
	$= 1.221 \times 10^{19}\,{\rm GeV}$
Gravitational constant	$G = 6.674 \times 10^{-5}\,{\rm cm^3/kg/s^2}$
	$= 6.709 \times 10^{-39}\,{\rm GeV^{-2}}$
Proton mass	$m_p = 1.6726 \times 10^{-27}\,{\rm kg}$
	$= 0.93827\,{\rm GeV}$
Neutron mass	$m_n = 1.6749 \times 10^{-27}\,{\rm kg}$
	$= 0.93957\,{\rm GeV}$
Electron mass	$m_e = 9.109 \times 10^{-31}\,{\rm kg}$
	$= 0.000511\,{\rm GeV}$
Neutral pion mass	$m_{\pi^0} = 0.1350\,{\rm GeV}$
Charged pion mass	$m_{\pi^\pm} = 0.1396\,{\rm GeV}$
$U(1)$ gauge coupling	$g_1 = 0.357$
$SU(2)$ gauge coupling	$g_2 = 0.652$
$SU(3)$ (strong) gauge coupling	$g_3 \equiv g_s = 1.181$ (evaluated at m_Z)
Weinberg (weak mixing) angle	$\sin^2\theta_W \equiv \frac{g_1^2}{g_1^2 + g_2^2} = 0.231$
Fine structure constant	$\alpha \equiv \frac{e^2}{4\pi} \equiv \frac{g_2^2 \sin^2\theta_W}{4\pi} = \frac{1}{137.0}$
Fermi coupling constant	$G_F \equiv \frac{\sqrt{2} g_2^2}{8 m_W^2} = 1.1664 \times 10^{-5}\,{\rm GeV^{-2}}$
Higgs vacuum expectation value (vev)	$v \equiv \left(\sqrt{2} G_F\right)^{-1/2} = 246\,{\rm GeV}$
Thomson cross section	$\sigma_t = \frac{8\pi\alpha^2}{3 m_e^2} = 6.652 \times 10^{-25}\,{\rm cm^2}$
	$= 1708\,{\rm GeV^{-2}}$

Note that these tables list only our best current determination of these quantities, and not any error bars or other uncertainties associated with these measurements.

Our cosmic timeline

Event	Time	Temperature	Redshift
Big Bang singularity (???)	0	∞	—
Planck era (???)	$\sim 10^{-43}$ s	$\sim 10^{19}$ GeV	—
Grand unified era (???)	$\lesssim 10^{-36}$ s	$\gtrsim 10^{15}$ GeV	—
Inflation (?)	$\sim 10^{-34}$ s (?)	—	?
Baryogenesis (?)	?	?	?
Dark matter freeze out (?)	?	?	?
Electroweak phase transition	$\sim 10^{-11}$ s	~ 200 GeV	$\sim 3 \times 10^{15}$
QCD phase transition	$\sim 10^{-5}$ s	~ 0.3 GeV	$\sim 3 \times 10^{12}$
Neutrino decoupling	~ 1 s	~ 1 MeV	$\sim 6 \times 10^{9}$
Proton-neutron freeze out	~ 1 s	~ 1 MeV	$\sim 6 \times 10^{9}$
Deuterium and helium fusion	$\sim 1-300$ s	$\sim 1-0.07$ MeV	$\sim (6-0.3) \times 10^{9}$
Free neutrons decay	~ 880 s	~ 30 keV	$\sim 2 \times 10^{5}$
Matter domination begins	$\approx 51{,}100$ yr	≈ 0.8 eV	≈ 3402
Recombination and photon decoupling	$\sim 380{,}000$ yr	~ 0.3 eV	~ 1090
First stars and galaxies	$\sim 50-200$ Myr	$\sim 0.01-0.005$ eV	$\sim 40-20$
Reionization	~ 0.5 Gyr	~ 0.002 eV	~ 8
Dark energy domination begins	~ 10 Gyr	~ 0.0003 eV	~ 0.3
Today	≈ 13.8 Gyr	≈ 0.000235 eV	0

Elementary particles

Leptons (spin 1/2)

	Generation	Charge	Mass (GeV)	Lifetime (s)	Y_{f_L}	Y_{f_R}	T_{f_L}	T_{f_R}
e (electron)	1	-1	0.000511	∞	-1	-2	1/2	0
ν_e (e neutrino)	1	0	$\lesssim 10^{-10}$	∞	-1	—	1/2	—
μ (muon)	2	-1	0.106	2.2×10^{-6}	-1	-2	1/2	0
ν_μ (μ neutrino)	2	0	$\lesssim 10^{-10}$	∞	-1	—	1/2	—
τ (tau)	3	-1	1.78	2.9×10^{-13}	-1	-2	1/2	0
ν_τ (τ neutrino)	3	0	$\lesssim 10^{-10}$	∞	-1	—	1/2	—

Quarks (spin 1/2)

	Generation	Charge	Mass (GeV)	Y_{f_L}	Y_{f_R}	T_{f_L}	T_{f_R}
d (down)	1	$-1/3$	$\sim 0.004-0.005$	1/3	$-2/3$	1/2	0
u (up)	1	$+2/3$	$\sim 0.002-0.003$	1/3	4/3	1/2	0
s (strange)	2	$-1/3$	$\sim 0.09-0.10$	1/3	$-2/3$	1/2	0
c (charm)	2	$+2/3$	1.27	1/3	4/3	1/2	0
b (bottom)	3	$-1/3$	4.18	1/3	$-2/3$	1/2	0
t (top)	3	$+2/3$	173	1/3	4/3	1/2	0

Bosons

	Charge	Color	Mass (GeV)	Lifetime (s)	Spin	Associated force
γ (photon)	0	0	0	∞	1	Electromagnetic
g (gluon)	0	color+anticolor	0	∞	1	Strong
$W^\pm$	± 1	0	80.4	3.1×10^{-25}	1	Weak
Z	0	0	91.2	2.6×10^{-25}	1	Weak
h (Higgs)	0	0	125.1	1.6×10^{-22}	0	Weak

Note that Y_{f_L}, Y_{f_R}, T_{f_L}, and T_{f_R} are quantities known as weak hypercharge and weak isospin, which characterize how the leptons and quarks interact with the electroweak gauge bosons, $W^\pm$ and Z.

Frequently Used Equations

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi\rho}{3M_{\rm Pl}^2} - \frac{\kappa}{a^2} + \frac{\Lambda}{3}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3M_{\rm Pl}^2}(\rho + 3p) + \frac{\Lambda}{3}$$

$$\rho = \frac{\pi^2 g_\star(T) T^4}{30}$$

$$T \propto \frac{1}{a\, g_{\star,S}^{1/3}}$$

$$n = \begin{cases} \dfrac{\zeta(3) g T^3}{\pi^2} & \text{(Bose-Einstein, } T \gg m\text{)} \\ \dfrac{3}{4}\dfrac{\zeta(3) g T^3}{\pi^2} & \text{(Fermi-Dirac, } T \gg m\text{)} \end{cases}$$

$$\rho = \begin{cases} \dfrac{\pi^2 g T^4}{30} & \text{(Bose-Einstein, } T \gg m\text{)} \\ \dfrac{7}{8}\dfrac{\pi^2 g T^4}{30} & \text{(Fermi-Dirac, } T \gg m\text{)} \end{cases}$$

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu)/T} \qquad (T \ll m)$$

$$\rho = g\, m^{5/2}\left(\frac{T}{2\pi}\right)^{3/2} e^{-(m-\mu)/T} \qquad (T \ll m)$$

Preface

The interface between cosmology, particle physics, and astrophysics is an exceptionally multidisciplinary area of research. A physicist investigating the nature of elementary particles, or studying condensed matter systems, might go their entire career without ever having any reason (beyond their own curiosity) to learn about or use the theory of general relativity. On similar grounds, an astrophysicist who specializes in the evolution of stars or planets might never have any clear use for quantum field theory. Particle-cosmologists, in contrast, often find themselves working on problems that call upon these and many other areas within physics and astrophysics. General relativity provides us with our foundation for understanding the expansion of our universe, and the role that gravity played in the evolution of large-scale structure. The many forms that energy takes in our universe, and how they have evolved over time, is dictated by quantum field theory and quantum statistical mechanics. Often in my career, my background—which is primarily in the area of theoretical particle physics—did not enable me to tackle the problems that I was most interested in exploring at the time. To pursue these problems, I had to look beyond my original training and learn more and dig deeper into other subjects, including thermodynamics, general relativity, galactic dynamics, magnetohydrodynamics, stellar and planetary astrophysics, nuclear physics, and many other areas of physics and astrophysics.

The multidisciplinary nature of this subject makes it a vibrant and exciting area to be working in. It also, however, presents substantial challenges for communicating many of the ideas that are involved. In the course of a single lifetime, no one could master all of the topics that might come up in this field of research. This means that scientists are often forced to specialize in one or more subfields, obtaining the necessary mastery in some areas, while leaving others comparatively neglected. It has been my experience that when listening to seminars at Fermilab's Cosmic Physics Center, or at the University of Chicago's Kavli Institute for Cosmological Physics, there is often a mismatch between what the speaker assumes their audience knows, and what is actually known by their audience. A scientist with a background in one area might assume knowledge about some topics that their audience does not possess, and a member of that audience might inadvertently return the favor one day when they are presenting their own research.

The same obstacles are faced by anyone writing a textbook about this subject. For most graduate-level textbooks, it is not difficult to identify the author's background—particle physicists, relativists, and other physicists and astronomers predictably assume very different things about their audiences, and place emphasis in very different areas. With

this book, I've attempted to write both for people with similar training to my own, and for those who might come from different academic backgrounds. Furthermore, realizing that everyone has gaps in their knowledge, especially when it comes to this exceptionally multidisciplinary subject, I've tried to explicitly fill in some of these gaps. This means that this book will be more pedagogical at times than most other graduate-level textbooks, especially in the first several chapters, as well as perhaps the final chapter. If you are a reader who already has a strong background in quantum field theory, you can probably skip Chapter 3, which is intended to provide an approachable introduction to particle physics, including Feynman diagrams, amplitudes, and the basic contours of the Standard Model. On similar grounds, if you have already taken a good course on general relativity, you probably won't gain very much by reading the first sections of Chapter 2, which introduce general relativity and the Friedmann equations. All of us have gaps in our knowledge. By being modular, it is my hope that this book will be able to address some of these gaps without holding back those readers who are already well versed in the material covered by a given section or chapter.

While this book could certainly be used in a graduate-level course in cosmology, particle astrophysics, or high-energy astrophysics, it could equally well be used as a reference for researchers in those fields. I originally thought I'd call this book either "Astrophysics for Particle Physicists" or "Particle Physics for Astronomers," but had a hard time picking between these two options. Whichever category you might fall into, I hope you find this book useful.

What this book is *not* is a substitute for a good textbook on cosmology. Most graduate-level cosmology textbooks, including those by Dodelson, Weinberg, and Baumann, discuss quite deeply the quantitative evolution of density perturbations, a subject which is important to cosmology and which is discussed in this book only in passing. To make room for a detailed treatment of this topic, I would have to remove chapters on other subjects which are not covered in most traditional books on cosmology. So although this book should probably not serve as a replacement for your standard cosmology reference, it aspires to be a complementary and valuable addition to your textbook library.

I would like to vigorously thank my friends and colleagues for their help with this project, including Nikita Blinov, Gordan Krnjaic, Alex Drlica-Wagner, Nick Gnedin, Austin Joyce, Pedro Machado, Dan Baxter, Ilias Cholis, Sam McDermott, Asher Berlin, and Elena Pinetti. I'd also like to thank the many students at the University of Chicago whose comments and suggestions have helped me to refine and improve this book, including Aurora Ireland, Rayne Liu, Ian Holst, Celeste Keith, Alex Hanselman, David Zegeye, Paul Chichura, Fei Xu, Madison Brady, Shreya Sutariya, Kaeli Hughes, Nina Coyle, Abby Lee, Keyer Thyme, David Martin, John Zhang, Austin Stover, Rachel Scrandis, and Gabe Hoshino. I'm forever indebted.

I

Preliminaries

1

Our Universe

In this first chapter, I outline our universe's history as we currently understand it, and describe the main observations and measurements that we have used to deduce and support these conclusions. I also provide a brief description of some of the open questions and puzzles which motivate much of the research in modern cosmology. Readers with a strong background in both cosmology and particle physics can feel free to skip this chapter.

A little over a century ago, there was nothing that one could call a science of cosmology. On the observational side, it was being actively debated until the 1920s whether the collection of stars known as the Milky Way constituted the entirety of our universe, or was instead just one of many "island universes," or what we would now call galaxies [1]. In terms of theory, physicists had developed a framework to describe how objects move through space, but those theories were fundamentally incapable of asking how space itself might change or evolve with time. To begin to meaningfully ask questions such as these, one had to wait until November of 1915, when Einstein completed and presented his general theory of relativity [2].

I would argue that observational cosmology really began with the work of Henrietta Leavitt [3], Vesto Slipher [4], Edwin Hubble, and Milton Humason [5], whose observations collectively demonstrated not only that many long-observed nebulae are in fact galaxies, independent of the Milky Way, but also that nearly all of these objects are moving away from our location in the universe. Furthermore, these galaxies are each receding with velocities that are approximately proportional to their distance from us, thus demonstrating that our universe is expanding with time.

Taken at face value, the fact that our universe is expanding also implies that it must be changing and evolving. Although some astronomers attempted to accommodate the expansion of space within the context of steady state cosmological models [6, 7], such efforts became only more contrived as the evidence accumulated. As early as the 1950s, observations had begun to support the fact that our universe was hotter and denser in the past than it is today. With the discovery of the cosmic microwave background in 1964, the so-called "Hot Big Bang" theory became the leading paradigm for our universe's history and evolution. In the decades that followed, the fact that our universe evolved from a hot and dense state gradually became a matter of scientific consensus.

To understand the earliest eras of our universe's history, it is essential to know how matter and energy behave under extremely hot and dense conditions. The first few hundred

thousand years after the Big Bang played out in large part according to the principles of atomic physics, while the first few seconds and minutes were dictated by the laws that govern the interactions of atomic nuclei. Modern particle accelerators allow us to study the behavior of particles at temperatures as high as roughly a TeV, probing times as early as a trillionth of a second after the Big Bang.

In these first moments, the seeds of what would become our universe were planted. The particles that make up the dark matter were created during this primordial era, along with the neutrinos, electrons, and nuclei that occupy our universe today. During this earliest period of time, significant quantities of matter somehow survived, while antimatter did not. By the end of inflation, density perturbations were already in place that would ultimately lead to the formation of our universe's large-scale structure. We cannot claim to understand our universe without understanding this brief but critically formative moment of time.

1.1 A brief history of our universe

As with many historical timelines, the farther cosmologists look into the past, the less certain they are about the events that unfolded then. Without any way to observe the earliest moments, or to study the conditions under which they existed, we are forced to blindly extrapolate beyond the regimes that are currently observationally or experimentally accessible to us. This has led many to speculate about a variety of possible events that may have taken place in the very early universe, including the following:

- **The Big Bang singularity** ($t = 0,\ T = \infty$): Perhaps our universe began as a spacetime singularity of infinite temperature and density, prior to which neither space nor time existed.
- **The Planck era** ($t \sim 10^{-43}$ s, $T \sim 10^{19}$ GeV): At temperatures near the Planck scale, the effects of quantum gravity are expected to have been in full effect, perhaps involving radical departures from the laws of physics as we currently understand them.
- **The era of grand unification** ($t \lesssim 10^{-36}$ s, $T \gtrsim 10^{15}$ GeV): In grand unified theories (GUTs), the three forces of the Standard Model are each part of a singular force (and symmetry group) that was broken as the universe cooled. During the grand unified era, this symmetry remained unbroken, blurring the distinction between these three forces, and between quarks and leptons.
- **The inflationary era** ($t = ?,\ T = ?$): The flatness and uniformity of our universe suggest that our universe may have undergone a period of exponential expansion during its very early history. While there now exists significant evidence for inflation, much remains unknown about this era and the physics behind it.

As our universe expanded, it eventually cooled to temperatures that we can study using particle accelerators. From the data collected at the Large Hadron Collider (LHC), the Tevatron, the Large Electron-Positron (LEP) collider, and other such experiments, we can deduce much about our universe's likely state during its first trillionth of a second, including the following:

- **The quark-gluon plasma** ($t \lesssim 10^{-5}$ s, $T \gtrsim 0.3$ GeV): During this era, our universe was filled with a dense plasma of quarks and gluons, containing all of the known species of particles.

- **The electroweak phase transition** ($t \sim 10^{-11}$ s, $T \sim 200$ GeV): Prior to this time, electroweak symmetry remained unbroken, and the known gauge bosons, quarks, and leptons were each massless. During this transition, these particles acquired masses through the Higgs mechanism, and the electromagnetic and weak forces took on their familiar forms.
- **The QCD phase transition** ($t \sim 10^{-5}$ s, $T \sim 0.3$ GeV): During this transition, the quarks and gluons became bound together into mesons and hadrons, resulting from the evolving strength of the strong force.

Although particle accelerators and other laboratory experiments allow us to speculate about our universe's first $\sim 10^{-12}$ seconds in a way that is reasonably well informed, significant events may have taken place during this period that accelerators are not yet capable of revealing to us. In particular, particles that are only feebly coupled to the Standard Model may have been very consequential during this period of time, while also being very difficult to produce and study in accelerators. Similarly, particles that were produced at temperatures greater than those probed by the Large Hadron Collider could have been present during this period of time, substantially impacting how our early universe evolved.

Moving forward in time, we find ourselves on much firmer ground. Beginning roughly a second after the Big Bang, we have both laboratory data and cosmological observations to inform our understanding of cosmic history. Through measurements of the primordial light element abundances, we can be quite confident that the following events took place in the early universe:

- **Neutrino decoupling** ($t \sim 1$ s, $T \sim 1$ MeV): Before our universe had become even one second old, its density had become low enough to allow it to become transparent to neutrinos. These particles have been traveling unimpeded ever since, and now constitute what is known as the cosmic neutrino background.
- **Proton-neutron freeze out** ($t \sim 1$ s, $T \sim 1$ MeV): Prior to this transition, protons could be converted into neutrons and vice versa through weak interactions. As our universe expanded, these processes became inefficient, and the abundances of protons and neutrons each became fixed (until the neutrons began to decay).
- **Deuterium and helium fusion** ($t \sim 1{-}300$ s, $T \sim 1{-}0.07$ MeV): During this era of Big Bang nucleosynthesis, free protons and neutrons fused together to form deuterons, followed by ^{3}He, ^{3}H, and ^{4}He. Within only a few minutes, approximately a quarter of all nucleons were bound within helium nuclei. A much smaller proportion of the nucleons were fused into the nuclei of lithium and beryllium.
- **The decay of the free neutrons** ($t \sim 880$ s, $T \sim 30$ keV): As the age of the universe became comparable to the neutron's lifetime, these particles gradually disappeared from our universe. Eventually, unstable nuclei such as ^{7}Be and ^{3}H followed suit.
- **Matter domination begins** ($t \sim 51{,}000$ yr, $T \sim 0.8$ eV): As photons and neutrinos lost their energy to the expansion of space, matter (including both baryons and dark matter) came to make up an increasingly large fraction of our universe's energy density, ultimately constituting the majority of the total energy.

It would greatly surprise me to learn that these events did not take place as described above. Our observational handles on this era are strong enough to give us considerable confidence in this overall picture. As we move further forward in time, this becomes only more true. From the era of photon decoupling onward, we have vast quantities of diverse

observational information to rely on. We can be approximately as certain about this period of our universe's history as the scientific method allows us to be about anything.

- **The formation of atoms** ($t \sim 380{,}000\,\text{yr}$, $T \sim 0.3\,\text{eV}$): As our universe cooled to temperatures well below the binding energies of hydrogen and helium, the overwhelming majority of the electrons became bound to these nuclei, forming the first electrically neutral atoms.
- **Photon decoupling** ($t \sim 380{,}000\,\text{yr}$, $T \sim 0.3\,\text{eV}$): As the plasma of charged particles that filled our universe transformed into a gas of neutral atoms, space became transparent to light for the first time. The photons that decoupled from the plasma during this era have been traveling through our universe ever since, and make up today's cosmic microwave background.
- **First stars and galaxies** ($t \sim 50{-}200\,\text{Myr}$, $T \sim 0.01{-}0.005\,\text{eV}$): Star formation began during this era, followed shortly thereafter by the formation of our universe's first galaxies.
- **Reionization** ($t \sim 0.5\,\text{Gyr}$, $T \sim 0.002\,\text{eV}$): Within half a billion years or so, energetic photons produced by the first stars and galaxies began to disassociate electrons from nuclei, returning our universe to a nearly fully ionized state.
- **Dark energy domination begins** ($t \sim 10\,\text{Gyr}$, $T \sim 0.0003\,\text{eV}$): For the past few billion years, the density of dark energy has exceeded that of matter, propelling our universe toward a state of exponential expansion.
- **Today** ($t \simeq 13.8\,\text{Gyr}$, $T \simeq 0.000235\,\text{eV}$): All of human history and culture plays out.

1.2 The empirical pillars of modern cosmology

On the opening page of Kolb and Turner's classic text, *The Early Universe*, the authors write, "Astronomy is a data-starved science. Cosmology is even more so." While this was a fair assessment at the time (circa 1990), it is certainly not the case today. Over the past few decades, cosmology has grown into an observationally rich science, with the ability to draw upon a huge body of diverse and precise data. In my view, it is this data that has made the present era of modern cosmology so exciting and vibrant. In this section, I'll summarize some of the most important portions of this data and what they tell us about our universe and its evolution.

1.2.1 The expansion rate

Although astronomers have been measuring the expansion rate of our universe since the time of Hubble and Humason, modern cosmology has dramatically improved such measurements in terms of both precision and the range of time (and redshift) across which such measurements can be made. When I took an undergraduate cosmology course in the late 1990s, my textbook reported that the current rate of Hubble expansion was probably somewhere in the range of $H_0 \sim 50{-}100\,\text{km/s/Mpc}$. Today, measurements of the cosmic microwave background indicate that $H_0 \simeq 67.4 \pm 0.5\,\text{km/s/Mpc}$ [8]. Even taking into account the fact that other measurement techniques currently seem to prefer values in the range of $H_0 \sim 72{-}76\,\text{km/s/Mpc}$ [9, 10, 11], it is clear that enormous progress has been made over the past two decades. Furthermore, while a variety of techniques have long been used to measure the local rate of Hubble expansion, cosmologists have over the past few decades begun to use objects such as Type Ia supernovae to measure how this

rate has evolved over most of our universe's history (i.e., out to redshifts of $z \gtrsim 2$). These measurements have had a great impact on our understanding of the composition of our universe, in particular, in revealing the presence of a significant density of dark energy.

1.2.2 The cosmic microwave background

Since its discovery in 1964, the cosmic microwave background (CMB) has been central to the field of observational cosmology. This collection of photons provides us with a detailed description of the state of our universe at the time of recombination, when electrons and nuclei formed the first electrically neutral atoms 380,000 years after the Big Bang. As a result of this transition from a plasma of charged particles to a gas of electrically neutral atoms, the photons decoupled from the matter, and have been more or less freely propagating through our universe ever since. The measured temperature anisotropies of the CMB tell us about how matter and energy were distributed throughout our universe at the time of recombination, as well as about the distribution of intervening matter. By scrutinizing the detailed properties of the CMB, cosmologists have been able to precisely determine the abundances of baryons, dark matter, and neutrinos in our universe, as well as the geometry of our universe itself.

1.2.3 The light element abundances

Although the nuclei of the heavier atomic species originated in stars, those of the lightest few elements (hydrogen, deuterium, helium, and lithium) were forged largely through nuclear fusion in the first seconds and minutes after the Big Bang. During this period of Big Bang nucleosynthesis (BBN), the temperature of our universe was hot enough to facilitate rapid fusion, $T \sim 1 - 0.01$ MeV, binding many of the free protons and neutrons together through the strong force. Measurements of the primordial helium-to-hydrogen and deuterium-to-hydrogen ratios each agree with the predictions of the Big Bang theory, and provide us with a measurement of the energy density and expansion rate of our universe as early as ~ 1 second after the Big Bang.

1.2.4 Large-scale structure

Over time, gravity causes regions that contain a greater than average density to collapse, forming increasingly dense structures. By studying the distribution of galaxies and galaxy clusters in our universe, we can infer facts about the expansion history of our universe, and the nature of dark matter. In particular, since the 1980s, observations of large-scale structure have made it clear that our universe contains large quantities of cold (or warm) dark matter. Such observations also provide us with evidence that dark energy has come to dominate our universe's energy density at late times.

1.2.5 Laboratory measurements

To reliably understand and interpret a given set of astronomical observations, one generally must possess knowledge of the laws of physics that are involved. For example, if physicists and chemists had not discovered and measured the spectral lines associated with various atomic species, cosmologists would not have been able to use redshift as a measure of distance. Similarly, if scientists had not performed laboratory experiments to study the interactions of nuclei at high temperatures, it is unlikely that astronomers would have learned much from their measurements of the primordial element abundances. In the current era, particle accelerators provide us with a foundation to understand our universe's first fraction of a second. In particular, by colliding pairs of protons together with 13.6 TeV

of energy, the Large Hadron Collider allows us to study the forms of matter and energy that populated our universe as early as a trillionth of a second after the Big Bang.

1.3 Cosmic concordance

It is remarkable how well the standard Big Bang theory has held up to observational scrutiny over the past several decades. Forty years ago, no large-scale galaxy survey had been completed, and no anisotropies had been detected in the CMB. And yet, as the avalanche of modern cosmological data subsequently accumulated, the same underlying cosmological model remained in place and continued to be supported by the data. The many precise measurements that were carried out during these decades certainly allowed us to determine the values of many of the parameters of the standard Big Bang theory, but they did not force us to discard or alter this model in any substantial way. With only a handful of free parameters, this long-standing model remains to this day capable of describing the vast collection of data that has been accumulated over this period of time. This standard theory, including the presence of dark energy and cold dark matter, has become known as the ΛCDM model.

Many of the facts pertaining to our universe and its history have now been independently determined using multiple probes. The density of matter, for example, has been determined with measurements of galaxies and galaxy clusters, as well as using the temperature anisotropies of the CMB. Furthermore, starting with the distribution of cold dark matter at the time of recombination as inferred from the CMB, we can calculate how this matter should be distributed in our universe today, and this prediction agrees well with the observed large-scale structure of our universe. On similar grounds, we have determined the density of dark energy by measuring the evolution of our universe's expansion rate, by measuring the evolution of large-scale structure, and by studying the CMB's temperature anisotropies.

The most restrictive constraints on our universe's evolution prior to the time of recombination come from measurements of the light element abundances. These data confirm that our universe expanded at approximately the predicted rate from $\sim$1 second after the Big Bang (corresponding to $T \sim$ MeV) onward, which in turn implies that it did not contain any large abundances of matter or energy beyond those described by the standard ΛCDM model.

As for understanding the era prior to BBN, we have essentially no direct observational probes to rely on. Instead, we use what we have learned from particle accelerators to deduce as much as we can about how this era likely played out. There are many ways, however, that such inferences could be unreliable. First of all, there could exist particle species that interact too feebly with the particles of the Standard Model to be detected at existing particle accelerators, while still being produced in significant abundances in the early universe. Second, even the Large Hadron Collider cannot probe the kinds of interactions that took place during our universe's first $\sim 10^{-12}$ seconds (corresponding to $T \gtrsim$ TeV). Any attempt we might make to describe this first trillionth of a second must, therefore, rely on significant extrapolations, leaving us with little reason to be confident in the predictions of our existing theories.

1.4 Open questions and outstanding puzzles

If the questions raised by the science of cosmology had all been answered, I would not have bothered to write this book. In fact, the field of particle cosmology is a vibrant one in large part because of the many important and intriguing puzzles that we have, so far,

failed to resolve. These open questions provide us with motivation to consider new models and theories, and to conduct new experiments and observations. Unanswered questions are the lifeblood of any healthy scientific community.

Below is a list of the questions and puzzles that at this time have not been conclusively addressed by our best current theories (general relativity and the Standard Model of particle physics).

- **Dark matter**: There is a near-consensus among cosmologists that most of the matter in our universe does not consist of atoms, but of something that does not appreciably radiate, reflect, or absorb light. It is not, however, yet known what this dark matter actually is. The existence of dark matter requires physics beyond the Standard Model, and almost certainly originated in the early universe, prior to BBN.
- **The matter-antimatter asymmetry**: The fact that our universe contains a significant abundance of baryonic matter but almost no antimatter also requires physics beyond the Standard Model. To generate this matter-antimatter asymmetry, there must exist exotic forms of matter that were present but were not in equilibrium in the early universe, whose interactions violate the conservation of both baryon number and charge-parity (CP) symmetry.
- **Dark energy**: Although general relativity allows for the possibility of vacuum energy in the form of the cosmological constant, quantum field theory leads us to expect that the density of this energy should be vastly larger than is observed in our universe. This apparent problem has led many cosmologists to explore dynamical mechanisms for dark energy, or to consider scenarios involving anthropic selection effects.
- **The flatness and horizon problems**: In the original form of the Big Bang theory, the degree to which the universe is curved increases as space expands. It thus requires highly fine-tuned initial conditions to explain the fact that our universe is approximately flat today. Furthermore, we observe regions of space that appear to have never been in causal contact, and yet are at nearly identical temperatures. These twin puzzles have motivated cosmologists to propose inflationary scenarios in which space expanded exponentially shortly after the Big Bang, driving the overall curvature toward zero and providing a mechanism to explain how the entire observable universe was once in causal contact.
- **The existence of neutrino masses**: Unlike in the case of other fermions, there is no mechanism in the Standard Model to generate masses for the neutrinos. The empirical fact that these particles have small masses requires new physics.
- **The hierarchy problem**: In the Standard Model, quantum corrections are expected to drive the mass of the Higgs boson to a very high value, well above that measured at the Large Hadron Collider. In lieu of an extreme fine-tuning of parameters, the relatively small value of the Higgs mass would appear to require physics beyond that of the Standard Model, typically involving new particles that are not much heavier than the TeV scale. Supersymmetry is the best known example of new physics that could resolve this problem.
- **The strong CP problem**: Despite the fact that the structure of QCD allows for the combined symmetries of charge and parity (CP) to be violated in the strong interactions, this has never been observed. To explain this without requiring a large degree of fine-tuning, dynamical mechanisms have been proposed that can drive the CP violating interactions to zero. The most well-known example of this is the

Peccei-Quinn mechanism, which leads to the prediction of a dark matter candidate in the form of the axion.

- **Quantum gravity**: Despite the incredible empirical successes of general relativity, it is a classical theory that must break down on scales shorter than the Planck length, where quantum effects become important. Unlike the three forces described by the Standard Model, gravity is nonrenormalizable and cannot be self-consistently quantized using the same approach. The question of how to reconcile general relativity with quantum field theory is perhaps the single most significant open question in all of contemporary physics.

Additional reading

For readers who are interested in the history of cosmology, I enthusiastically recommend *Cosmology's Century* by Peebles. At the risk of coming across as overly self-promoting, I'll also suggest my own book, *At the Edge of Time*. On the particle physics side, I have particularly enjoyed Weinberg's *The Discovery of Subatomic Particles* and *The Second Creation* by Crease and Mann.

2

General Relativity and the Friedmann Equations

In this chapter, I introduce some of the concepts from general relativity that are most important and useful in cosmology, including the Friedmann-Lemaître-Robertson-Walker metric. I also derive the Friedmann equations, which describe how a universe expands or contracts with time. A reader who is familiar with general relativity and has already gone through this derivation could safely skip ahead to the second half of this chapter, where I describe several of the cosmological models that can arise from the Friedmann equations, and discuss the kinds of distances that are most commonly used in cosmology.

Central to the subject of cosmology is the fact that our universe evolves with time. Before Einstein's general theory of relativity, physicists had no way of even conceptualizing how this might occur. As we will see in this chapter, however, general relativity not only allows for the possibility that space can change, but essentially requires that it must expand or contract with time.

2.1 Spacetime and the metric

At the heart of the general theory of relativity are the Einstein equations (sometimes called the Einstein field equations):

$$R_{\mu\nu} - \frac{1}{2}\mathcal{R}g_{\mu\nu} = \frac{8\pi T_{\mu\nu}}{M_{\rm Pl}^2}. \tag{2.1}$$

The two terms on left-hand side of this equation collectively make up what is known as the Einstein tensor, which describes the curvature or geometry of spacetime. On the right-hand side is a tensor which describes the matter and other energy that is contained within that spacetime (more about tensors shortly).[1] To put this equation into words, the quantity and distribution of energy that is present determines the geometry of spacetime, which in turn

[1] Some readers might be used to seeing the right-hand side of Eq. 2.1 written in terms of the Newtonian gravitational constant, $8\pi G T_{\mu\nu}$. Throughout this book, I will instead present this quantity in terms of the Planck mass, which in natural units is $M_{\rm Pl} \equiv G^{-1/2}$.

determines how objects travel through that spacetime. As John Wheeler famously summed it up, "Spacetime tells matter how to move; matter tells spacetime how to curve."

Among the earliest applications of the Einstein equations were calculations of the orbits of planets (in particular, the rate of the precession of Mercury's perihelion), and the deflection of light by the sun and other massive bodies. As we will see, these equations can also be used to deduce how space evolves over time, making it of central importance to the subject of cosmology.

The Einstein tensor, which makes up the left-hand side of Eq. 2.1, is constructed from the metric, $g_{\mu\nu}$, the Ricci tensor, $R_{\mu\nu}$, and the Ricci scalar, $\mathcal{R}$. Furthermore, $R_{\mu\nu}$ and $\mathcal{R}$ are each functions of the metric, making the entire Einstein tensor an expression of the geometry of spacetime. The right-hand side of the Einstein equations describes the distribution of energy in terms of the energy-momentum tensor, $T_{\mu\nu}$. If you are not familiar with these symbols, no need to worry. I'll go through them one by one in the discussion below.

When Einstein completed his general theory of relativity, he didn't seem to have had any cosmological applications in mind. But within a year or so, he had started to think about ways to apply his field equations to the universe as a whole. To make this problem tractable, he chose to adopt two key assumptions: namely, that the universe is both homogeneous and isotropic on very large scales. In other words, he assumed that when averaged over very large volumes of space, the density, composition, and curvature of the universe are approximately the same everywhere and in all directions. While these assumptions were originally adopted for the sake of mathematical convenience, they have since been confirmed by numerous observations—the cow, in this case, has turned out to be remarkably spherical! Although our universe is not at all homogeneous or isotropic on the modest scales of solar systems or galaxies, modern surveys indicate that it *is* approximately uniform on scales larger than ~ 100 Mpc or so. For example, all $(100\ \text{Mpc})^3$ regions of space in the present universe contain approximately the same number of galaxies.

Given the assumed requirements of homogeneity and isotropy, we can ask what spatial metrics one could potentially consider. An obvious choice is that of an ordinary flat, or Euclidean, metric:

$$\begin{aligned} dl^2 &= dx^2 + dy^2 + dz^2 \\ &= dr^2 + r^2[d\theta^2 + \sin^2\theta\, d\phi^2], \end{aligned} \tag{2.2}$$

where the first and second lines are written in Cartesian and spherical coordinates, respectively.

As a second possibility, we could also consider a metric with uniform positive curvature:

$$dl^2 = dr^2 + R^2 \sin^2\left(\frac{r}{R}\right)[d\theta^2 + \sin^2\theta\, d\phi^2]. \tag{2.3}$$

This is the three-dimensional analog of the metric that describes the two-dimensional surface of a sphere. In such a space, parallel lines converge, the angles of triangles sum to more than 180°, and the circumference of a circle is always less than 2π times its radius. The magnitude of the spatial curvature is determined by the value of the parameter R. Notice that in the $r \ll R$ limit, this metric reduces to that of the Euclidean case.

The third metric that satisfies these requirements is one with uniform negative curvature. This is arrived at by taking the positive curvature case and making the replacement $R^2 \to -R^2$. Applying Euler's formula, this substitution results in the following:

$$
\begin{aligned}
dl^2 &= dr^2 - R^2 \sin^2\left(\frac{r}{\sqrt{-R^2}}\right)[d\theta^2 + \sin^2\theta\, d\phi^2] \qquad (2.4)\\
&= dr^2 - R^2\left[\frac{e^{ir/\sqrt{-R^2}} - e^{-ir/\sqrt{-R^2}}}{2i}\right]^2 [d\theta^2 + \sin^2\theta\, d\phi^2]\\
&= dr^2 + R^2\left[\frac{e^{r/R} - e^{-r/R}}{2}\right]^2 [d\theta^2 + \sin^2\theta\, d\phi^2]\\
&= dr^2 + R^2 \sinh^2\left(\frac{r}{R}\right)[d\theta^2 + \sin^2\theta\, d\phi^2].
\end{aligned}
$$

Most people (including myself) find negatively curved space to be more difficult to visualize than flat or positively curved space. In a negatively curved space, parallel lines diverge from one another, the angles of triangles add up to less than 180°, and the circumference of a circle is always greater than 2π times its radius.

Up to an overall multiplicative factor, these three cases are the *only* three-dimensional metrics that are consistent with the assumptions of homogeneity and isotropy. It is useful to recast these expressions by making the following change of coordinates:

$$
\begin{aligned}
R\sin(r/R) &\to r \qquad \text{(positive curvature)}, \qquad (2.5)\\
R\sinh(r/R) &\to r \qquad \text{(negative curvature)},
\end{aligned}
$$

which leads to the following, more useful, expression for the spatial metric:

$$
dl^2 = \frac{dr^2}{1 - kr^2/R^2} + r^2[d\theta^2 + \sin^2\theta\, d\phi^2], \qquad (2.6)
$$

where $k = +1, 0, -1$ for the cases of positive, zero, and negative curvature, respectively.

Of course, we are not really interested in the spatial metric, but in the metric that describes intervals in *spacetime*. In the context of special relativity, we would write the separation between two events in spacetime in terms of the Minkowski metric:

$$
ds^2 = dt^2 - dr^2 - r^2[d\theta^2 + \sin^2\theta\, d\phi^2]. \qquad (2.7)
$$

In general relativity, we must further allow for the possibility of spacetime curvature, as well as for variations of the metric with time. We do this by combining the Minkowski metric with the spatial metric given in Eq. 2.6, weighted by a scale factor, $a(t)$, which describes the relative expansion or contraction of space with respect to time:

$$
ds^2 = dt^2 - a^2(t)\left[\frac{dr^2}{1 - \kappa r^2} + r^2(d\theta^2 + \sin^2\theta\, d\phi^2)\right]. \qquad (2.8)
$$

While here I have explicitly noted the time dependance of the scale factor, this will usually be implied from now on. The expression given in Eq. 2.8 is known as the Friedmann-Lemaître-Robertson-Walker metric, and it is the most general spacetime metric that is consistent with the assumptions of homogeneity and isotropy. Note that this expression is

left unchanged under the following substitutions:

$$\kappa \to \frac{\kappa}{|\kappa|}, \tag{2.9}$$
$$r \to \sqrt{|\kappa|}\, r,$$
$$a \to \frac{a}{\sqrt{|\kappa|}}.$$

Because of this, we can always choose to rescale r and a such that κ is exactly equal to $+1$, 0, or -1. Alternatively, one can define a such that it is equal to unity at the present time, in which case κ will not generally be an integer.

We can write the Friedmann-Lemaître-Robertson-Walker metric more compactly in terms of the 4×4 matrix known as the metric tensor, $g_{\mu\nu}$:

$$ds^2 = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu. \tag{2.10}$$

To be more explicit, the various components of the metric tensor are given by:

$$g_{00} = 1, \qquad g_{11} = -\frac{a^2}{1-\kappa r^2}, \qquad g_{22} = -a^2 r^2, \tag{2.11}$$
$$g_{33} = -a^2 r^2 \sin^2\theta, \qquad g_{\mu\nu} = 0 \quad (\text{for all } \mu \neq \nu).$$

2.2 Tensor notation in general relativity

In general relativity, there is a system of notation that is commonly used that, while intended to make it easier to keep track of the many tensors and vectors involved, may be unfamiliar to some readers. In this section, I will provide a brief summary of these conventions and how they are most commonly used.

When a symbol has as a subscript or superscript a Greek letter (μ, ν, α, β, etc.), such indices represent the four dimensions of spacetime, with the first dimension (0) being reserved for time, and the others (1, 2, 3) for space. In contrast, Latin indices (i, j, k, etc.) are used to denote the three spatial dimensions (1, 2, 3) without time. Whenever the same index appears twice within the same term as both a subscript and a superscript, it is automatically implied that a summation is performed over that index. For example, the metric tensor given in Eq. 2.10 can be written as:

$$ds^2 = \sum_{\mu=0}^{3} \sum_{\nu=0}^{3} g_{\mu\nu} dx^\mu dx^\nu \tag{2.12}$$
$$= g_{\mu\nu} dx^\mu dx^\nu.$$

The quantities dx^μ and dx^ν are vectors, each with four components. In addition to vectors, objects known as one-forms are often used in general relativity. One-forms are similar to vectors, but are written with a lowered index, A_μ. Vectors and one-forms are related to one

another through the metric tensor as follows:

$$A^{\mu} = g^{\mu\nu} A_{\nu}, \tag{2.13}$$
$$A_{\mu} = g_{\mu\nu} A^{\nu},$$

where $g^{\mu\nu}$ is the inverse of the metric tensor, $g_{\mu\nu}$.

The metric tensor can also be used to raise or lower the indices on other tensors. For example:

$$A^{\mu\nu} = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} g^{\mu\alpha} g^{\nu\beta} A_{\alpha\beta} \tag{2.14}$$
$$= g^{\mu\alpha} g^{\nu\beta} A_{\alpha\beta}.$$

Applying Eq. 2.14 to the case of the metric tensor itself, $g^{\mu\nu} = g^{\mu\alpha} g^{\nu\beta} g_{\alpha\beta}$, we can deduce the following useful relation:

$$g^{\nu\beta} g_{\alpha\beta} = \delta^{\nu}_{\alpha}, \tag{2.15}$$

where δ^{ν}_{α} is the Kronecker delta, which is equal to 1 if $\nu = \alpha$, and to 0 otherwise. From this, we can see that $g^{\mu\nu}$ must be the inverse of $g_{\mu\nu}$.

2.3 The Friedmann equations

The Friedmann equations describe how the scale factor of a homogeneous and isotropic universe evolves with time, and they are among the most important equations in all of cosmology. In this section, we'll see how these equations can be derived from the Einstein equations.

In relativity, particles that experience no external forces move along trajectories known as geodesics. This is something of a generalization of Newton's First Law, which states that particles in the absence of forces move along straight lines. As a result of the curvature of spacetime, however, geodesics are not necessarily straight—at least not in the way that we thought about things being straight prior to Einstein. Instead, geodesics are described by the following equation:

$$\frac{d^2 x^{\mu}}{ds^2} + \Gamma^{\mu}_{\alpha\beta} \frac{dx^{\alpha}}{ds} \frac{dx^{\beta}}{ds} = 0, \tag{2.16}$$

where s is a parameter of motion (such as proper time), and $\Gamma^{\mu}_{\alpha\beta}$ are the Christoffel symbols, which can be derived from the metric tensor:

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2} g^{\mu\nu} \left[\frac{\partial g_{\alpha\nu}}{\partial x^{\beta}} + \frac{\partial g_{\beta\nu}}{\partial x^{\alpha}} - \frac{\partial g_{\alpha\beta}}{\partial x^{\nu}} \right]. \tag{2.17}$$

Since the Christoffel symbols are functions only of the metric tensor, it is the metric alone that determines the geodesics.

As we will see, we will need the Christoffel symbols to calculate how space expands or contracts with time. Let's start with the $\mu = 0, \alpha = \beta = 1$ component:

$$\begin{aligned}\Gamma^0_{11} &= \frac{1}{2}g^{0\nu}\left[\frac{\partial g_{1\nu}}{\partial x^1} + \frac{\partial g_{1\nu}}{\partial x^1} - \frac{\partial g_{11}}{\partial x^\nu}\right] \\ &= \frac{1}{2}g^{00}\left[\frac{\partial g_{10}}{\partial x^1} + \frac{\partial g_{10}}{\partial x^1} - \frac{\partial g_{11}}{\partial t}\right] \\ &= \frac{1}{2}\left[0 + 0 + \frac{2a\dot{a}}{1-\kappa r^2}\right] \\ &= \frac{a\dot{a}}{1-\kappa r^2},\end{aligned} \tag{2.18}$$

where dotted quantities refer to time derivatives, $\dot{a} \equiv da/dt$. Similarly, we can derive the other components of $\Gamma^\mu_{\alpha\beta}$ from Eq. 2.17 and the metric tensor:

$$\begin{aligned}&\Gamma^0_{22} = a\dot{a}r^2, && \Gamma^0_{33} = a\dot{a}r^2\sin^2\theta, \\ &\Gamma^1_{11} = \frac{\kappa r}{1-\kappa r^2}, && \Gamma^1_{01} = \Gamma^1_{10} = \Gamma^2_{02} = \Gamma^2_{20} = \Gamma^3_{03} = \Gamma^3_{30} = \frac{\dot{a}}{a}, \\ &\Gamma^1_{22} = -r(1-\kappa r^2), && \Gamma^1_{33} = -r(1-\kappa r^2)\sin^2\theta, \\ &\Gamma^2_{12} = \Gamma^2_{21} = \Gamma^3_{13} = \Gamma^3_{31} = \frac{1}{r}, && \Gamma^3_{23} = \Gamma^3_{32} = \cot\theta, \\ &\Gamma^2_{33} = -\sin\theta\cos\theta, &&\end{aligned} \tag{2.19}$$

with all other elements being equal to zero. With the Christoffel symbols in hand, we are now in a position to evaluate the Ricci tensor, which first appeared in Eq. 2.1, and is defined as follows:

$$R_{\mu\nu} \equiv \Gamma^\alpha_{\mu\nu,\alpha} - \Gamma^\alpha_{\mu\alpha,\nu} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{\mu\nu} - \Gamma^\alpha_{\beta\nu}\Gamma^\beta_{\mu\alpha}, \tag{2.20}$$

where the commas denote derivatives, such as $\Gamma^\alpha_{\mu\nu,\alpha} \equiv \partial\Gamma^\alpha_{\mu\nu}/\partial x^\alpha$.

Starting with R_{00}:

$$\begin{aligned}R_{00} &= \Gamma^\alpha_{00,\alpha} - \Gamma^\alpha_{0\alpha,0} + \Gamma^\alpha_{\beta\alpha}\Gamma^\beta_{00} - \Gamma^\alpha_{\beta0}\Gamma^\beta_{0\alpha} \\ &= 0 - \Gamma^1_{01,0} - \Gamma^2_{02,0} - \Gamma^3_{03,0} + 0 - \Gamma^1_{10}\Gamma^1_{01} - \Gamma^2_{20}\Gamma^2_{02} - \Gamma^3_{30}\Gamma^3_{03} \\ &= -3\frac{\partial}{\partial t}\left(\frac{\dot{a}}{a}\right) - 3\left(\frac{\dot{a}}{a}\right)^2 \\ &= -3\frac{\ddot{a}}{a}.\end{aligned} \tag{2.21}$$

And similarly,

$$R_{11} = -\frac{a\ddot{a} + 2\dot{a}^2 + 2\kappa}{1 - \kappa r^2}, \tag{2.22}$$

$$R_{22} = -r^2(a\ddot{a} + 2\dot{a}^2 + 2\kappa),$$

$$R_{33} = -r^2 \sin^2\theta(a\ddot{a} + 2\dot{a}^2 + 2\kappa),$$

and all other elements of $R_{\mu\nu}$ are zero.

The Ricci scalar, $\mathcal{R}$, is calculated by contracting the metric with the Ricci tensor:

$$\begin{aligned}\mathcal{R} &\equiv g^{\mu\nu}R_{\mu\nu} \\ &= -6\left(\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{\kappa}{a^2}\right).\end{aligned} \tag{2.23}$$

At this point, we have almost everything that we will need to derive the Friedmann equations from the Einstein equations. The only thing we are missing is the energy-momentum tensor, $T_{\mu\nu}$. Continuing under the assumption of homogeneity and isotropy, this quantity is most commonly written as follows:

$$T^{\mu}_{\nu} = \begin{pmatrix} \rho & 0 & 0 & 0 \\ 0 & -p & 0 & 0 \\ 0 & 0 & -p & 0 \\ 0 & 0 & 0 & -p \end{pmatrix}, \tag{2.24}$$

where ρ and p are the energy density and pressure of the uniform matter and energy contained within the universe. Since Eq. 2.1 is written in terms of $T_{\mu\nu}$ rather than T^{μ}_{ν}, we need to lower an index as follows:

$$\begin{aligned}T_{\mu\nu} &= g_{\mu\alpha}T^{\alpha}_{\nu} \\ &= g_{\mu 0}T^{0}_{\nu} + g_{\mu 1}T^{1}_{\nu} + g_{\mu 2}T^{2}_{\nu} + g_{\mu 3}T^{3}_{\nu}.\end{aligned} \tag{2.25}$$

Writing just the $(0, 0)$ component of Eq. 2.1, we have:

$$R_{00} - \frac{1}{2}\mathcal{R}g_{00} = \frac{8\pi T_{00}}{M_{\mathrm{Pl}}^2}. \tag{2.26}$$

After inserting our expressions for R_{00}, $\mathcal{R}$, g_{00}, and T_{00}, we arrive at the first Friedmann equation:

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = \frac{8\pi\rho}{3M_{\mathrm{Pl}}^2}. \tag{2.27}$$

Similarly, from any of the $(1, 1)$, $(2, 2)$, or $(3, 3)$ components, we can obtain the second Friedmann equation:

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = -\frac{8\pi p}{M_{\mathrm{Pl}}^2}. \tag{2.28}$$

Alternatively, we can use the first Friedmann equation to replace the terms containing $\dot{a}$ and κ in Eq. 2.28 with one involving ρ, allowing us to rewrite the second equation as follows:

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3M_{\rm Pl}^2}(\rho + 3p). \tag{2.29}$$

2.4 Matter, radiation, and pressure

In order to make use of the Friedmann equations, we will have to evaluate the energy density and pressure of various components of matter and radiation that do (or could) exist in our universe. Although I will discuss the subject of thermodynamics as applied to cosmology in much more detail in Chapter 4, here we will focus on a few particularly relevant points. Consider a gas of particles in equilibrium at a temperature T with no chemical potential. The energy density of this gas is given by

$$\begin{aligned}\rho &= \int \frac{dn}{d^3\vec{p}}\, E\, d^3\vec{p} \\ &= \frac{g}{2\pi^2}\int_m^\infty \frac{(E^2-m^2)^{1/2}}{e^{E/T}\pm 1} E^2 dE,\end{aligned} \tag{2.30}$$

where the integral in the first line is over all momentum, and g is the number of internal degrees of freedom of the particle species. The positive sign in the denominator corresponds to the case of fermions, which follow the Fermi-Dirac distribution, while the negative sign is for the case of bosons, which are described by the Bose-Einstein distribution. The analogous expression for the pressure of the gas is

$$p = \frac{g}{6\pi^2}\int_m^\infty \frac{(E^2-m^2)^{3/2}}{e^{E/T}\pm 1} dE. \tag{2.31}$$

Let's consider two limiting cases. First, if the particles are highly non-relativistic, $T \ll m$, then these expressions reduce to

$$\begin{aligned}\rho &= gm\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T}, \\ p &= gT\left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T},\end{aligned} \tag{2.32}$$

and therefore $p = \rho T/m \ll \rho$. When considering nonrelativistic matter, one usually takes the pressure to be equal to zero when evaluating the second Friedmann equation.

Next, if the particles are highly relativistic, $T \gg m$, then we instead have

$$\rho = \begin{cases} \dfrac{7}{8}\dfrac{\pi^2 g T^4}{30} & \text{(Fermi-Dirac)} \\ \dfrac{\pi^2 g T^4}{30} & \text{(Bose-Einstein)} \end{cases} \tag{2.33}$$

and

$$p = \begin{cases} \frac{7}{8}\frac{\pi^2 g T^4}{90} & \text{(Fermi-Dirac)} \\ \frac{\pi^2 g T^4}{90} & \text{(Bose-Einstein).} \end{cases} \tag{2.34}$$

For either fermions or bosons, we find $p = \rho/3$. Note that cosmologists usually refer to nonrelativistic matter simply as "matter" and to relativistic particles as "radiation," and I will do the same throughout this book. We can generalize the ratio of the pressure to the energy density by introducing the equation of state, w:

$$w \equiv \frac{\sum_i p_i}{\sum_i \rho_i}, \tag{2.35}$$

where the summation is performed over all of the forms of matter and radiation that are present. For a universe that is entirely dominated by matter, $w \approx 0$, whereas one containing mostly radiation has $w \approx 1/3$.

2.5 Static universes and the cosmological constant

When Einstein first began to contemplate the cosmological consequences of general relativity, he seems to have held a prejudice that our universe should be static and unchanging [12]. As we will see, this is not a particularly easy position to hold in light of what general relativity tells us about the nature of space and time.

In order for a universe to be static, as opposed to expanding or contracting, its scale factor must be constant, $\dot{a} = \ddot{a} = 0$. Making this substitution into Eqs. 2.27 and 2.29, we arrive at the following:

$$\frac{\kappa}{a^2} = \frac{8\pi\rho}{3M_{\rm Pl}^2} \tag{2.36}$$

and

$$\rho = -3p. \tag{2.37}$$

Thus, in order for a homogeneous and isotropic universe to be static, it must have positive curvature ($\kappa > 0$), since scale factors and energy densities are each uniquely positive. In addition, the pressure and energy density must be related by $\rho = -3p$, corresponding to $w = -1/3$. From the previous section, we know that any combination of matter and radiation should correspond to an equation of state between 0 and 1/3. Thus, if there is to be any chance of finding a static solution to the Einstein equations, the universe has to be dominated by something quite exotic, possessing a significant degree of negative pressure.

When faced with this problem, one might have expected Einstein to have acknowledged that his theory seems to insist that space should be either expanding or contracting. Instead, he remained determined to find a static solution, leading him to revisit the Einstein equations themselves and modify them to include an additional term:

$$R_{\mu\nu} - \frac{1}{2}\mathcal{R}g_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi T_{\mu\nu}}{M_{\rm Pl}^2}. \tag{2.38}$$

The newly included term, $-\Lambda g_{\mu\nu}$, is sometimes called the "cosmological term," and the symbol Λ represents a number known as the "cosmological constant." In modifying this equation, Einstein did not have many options to choose from. Technically speaking, the left-hand side of Eq. 2.38 is the only local, covariant, four-dimensional set of gravitational field equations that can be constructed from the metric and its first and second derivatives [13].

When the cosmological term is included, the resulting Friedmann equations become:

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} - \frac{\Lambda}{3} = \frac{8\pi\rho}{3M_{\rm Pl}^2}, \tag{2.39}$$

and

$$\frac{\ddot{a}}{a} - \frac{\Lambda}{3} = -\frac{4\pi}{3M_{\rm Pl}^2}(\rho + 3p). \tag{2.40}$$

There are two very different ways of interpreting the role of the cosmological term in the Einstein equations. One option is to write this term as part of the left-hand side of the Einstein equations (as I have in Eq. 2.38), and to view it as part of the Einstein tensor. From this perspective, by including the cosmological term, we have changed the way that the geometry of spacetime responds to energy. In other words, we have modified the nature of gravity. Alternatively, we could move the cosmological term to the right-hand side of the Einstein equations, and treat it as part of the energy-momentum tensor. In this context, one could view the cosmological constant as corresponding to a new component of energy in the universe, with its own contribution to the energy density and pressure. In particular, from the first Friedmann equation, we can see that the energy density associated with the cosmological constant is equal to $\rho_\Lambda = \Lambda M_{\rm Pl}^2/8\pi$. On similar grounds, from the second Friedmann equation, we can show that the pressure associated with the cosmological constant is given by $p_\Lambda = -\Lambda M_{\rm Pl}^2/8\pi$, corresponding to an equation of state of $w_\Lambda = -1$. Ah ha! This is exactly the negative pressure that Einstein was looking for in order to obtain a static cosmological solution!

Once again setting $\ddot{a} = \dot{a} = 0$, we now find that a static solution to the Friedmann equations *does* exist. Other than that associated with the cosmological constant, let's assume that the energy density of the universe is dominated by matter ($p = 0$). From the second Friedmann equation, we find that the value of the cosmological constant must be $\Lambda = 4\pi\rho/M_{\rm Pl}^2$ in this case, where ρ is the density of matter. Inserting this into the first Friedmann equation, we find that $\kappa/a^2 = 4\pi\rho/M_{\rm Pl}^2 = \Lambda$. This result tells us that in order for the universe to be static, spacetime must be positively curved, $\kappa > 0$. In a spacetime with uniform positive curvature, the universe wraps around on itself, like the surface of a sphere. Furthermore, this result also tells us about the degree of curvature that is required. Using the modern measured value for the average matter density, $\rho \approx 2.7 \times 10^{-27}$ kg/m^3, we find that one could circumnavigate Einstein's static universe by traveling a distance of $2\pi a/\sqrt{\kappa} = 2\pi c/\sqrt{4\pi\rho/M_{\rm Pl}^2} \sim 2 \times 10^{26}$ m ~ 6.5 Gpc. Given that it wasn't even clear at the time what (if anything) existed beyond the boundaries of the Milky Way, it's easy to see why Einstein did not immediately rule out such a possibility.

The real problem with Einstein's static cosmological solution is that it is unstable. If the energy density of the universe were even slightly greater or smaller than $\Lambda M_{\rm Pl}^2/4\pi$, the scale factor would begin to decrease or increase, respectively. A universe in a static configuration is like a pencil perfectly balanced on its tip—a static solution to the equations of motion does exist, but this solution does not describe a state that the universe will remain in for very long. Furthermore, the fact that our universe is not perfectly homogeneous and isotropic

will inevitably lead to regions that expand or contract with time. Thus there are no realistic static solutions to the equations of general relativity.

Ultimately, Einstein's preference for a static cosmological solution was overcome by Edwin Hubble and Milton Humason's discovery in 1929 [5] (made possible by the earlier work of Henrietta Leavitt [3] and Vesto Slipher [4]) that our universe is, in fact, expanding. Within a couple of years, Einstein had discarded his cosmological term, and had accepted these new observational results. Many decades later, the cosmological constant would be reintroduced into modern cosmology, playing the very different—but essential—role of what is now known as dark energy (see Chapter 14).

2.6 The evolution of energy

As the scale factor of the universe changes, so does the energy density of matter and energy. In a universe that was neither expanding nor contracting, we could express the combined conservation of energy and momentum simply as $\partial T^{\mu}_{\nu}/\partial x^{\mu} = 0$. In the presence of expansion or contraction, however, we instead have to equate the covariant derivative of T^{μ}_{ν} to zero:

$$\nabla_{\mu} T^{\mu}_{\nu} \equiv \frac{\partial T^{\mu}_{\nu}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu} T^{\alpha}_{\nu} - \Gamma^{\alpha}_{\nu\mu} T^{\mu}_{\alpha} = 0. \tag{2.41}$$

Focusing on the $\nu = 0$ piece of this equation, we have

$$\frac{\partial T^{\mu}_{0}}{\partial x^{\mu}} + \Gamma^{\mu}_{\alpha\mu} T^{\alpha}_{0} - \Gamma^{\alpha}_{0\mu} T^{\mu}_{\alpha} = 0. \tag{2.42}$$

Inputting the appropriate components of the energy-momentum tensor and the Christoffel symbols, this becomes

$$\frac{\partial \rho}{\partial t} + \frac{3\dot{a}(\rho + p)}{a} = 0. \tag{2.43}$$

This can be written in a more suggestive way by focusing on the comoving energy density, ρa^3:

$$\frac{\partial (\rho a^3)}{\partial t} = -3\dot{a} a^2 p. \tag{2.44}$$

In the case of nonrelativistic matter ($p = 0$), this equation says that the comoving energy remains constant, and thus $\rho \propto a^{-3}$. This is simply the consequence of geometric dilution; as the universe expands and its volume increases by a given factor, the energy density is reduced by the same factor.

If we instead consider the case of relativistic particles (i.e., radiation), we relate $p = \rho/3$, leading to the following result:

$$\begin{aligned} \frac{\partial (\rho a^4)}{\partial t} &= 4a^3 \dot{a}\rho + a^4 \frac{\partial \rho}{\partial t} \\ &= 4a^3 \dot{a}\rho - 4a^3 \dot{a}\rho \\ &= 0. \end{aligned} \tag{2.45}$$

So whereas matter is diluted by three powers of the scale factor, the energy density of radiation is reduced by four, $\rho \propto a^{-4}$. Three of these powers of the scale factor correspond to the same geometric dilution that is experienced by matter, while the additional factor of a is the result of cosmological redshift. As the scale factor increases, so does each particle's wavelength, $\lambda \propto a$. Since $E \propto \lambda^{-1}$ for relativistic particles, the expansion of space reduces the energy of such particles by a factor of a. This causes photons and other relativistic particles to reach earth with a longer wavelength than they were emitted with:

$$1 + z \equiv \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} = \frac{1}{a}, \tag{2.46}$$

where z is the redshift of the particle's source and a was the scale factor of the universe when the particle was emitted.

Generalizing this discussion, we can describe the evolution of an energy density with an arbitrary equation of state as follows:

$$\rho \propto a^{-3(1+w)}. \tag{2.47}$$

It can quickly be shown that this expression recovers the expected behavior for matter ($w = 0$) and radiation ($w = 1/3$). For the case of a cosmological constant ($w = -1$), we find the strange result that ρ_Λ remains constant as the universe expands or contracts. In other words, the expansion of the universe does not dilute or otherwise impact the energy density associated with the cosmological constant. For this reason, the cosmological constant is often thought of as the energy density that is intrinsically associated with the vacuum of our universe.

2.7 The expansion history of the universe

It is common to present the first Friedmann equation in terms of the expansion rate, $H \equiv \dot{a}/a$:

$$H^2 \equiv \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi\rho}{3M_{\text{Pl}}^2} - \frac{\kappa}{a^2}. \tag{2.48}$$

Although I have not explicitly included the cosmological constant in this expression, the energy density, ρ, can include a contribution associated with Λ, making this version of the first Friedmann equation quite general. Note that the current value of H is known as the Hubble constant, although it may seem strange that we call the current value of an evolving quantity a "constant." More generally, I'll refer to H as either the Hubble parameter or the Hubble rate, and to H_0 as the Hubble constant, defined as the current value of this quantity.

For a given value of H at a particular point in time, the curvature of space is related to the energy density. In particular, a universe that is flat ($\kappa = 0$) will contain an energy density that is equal to what is known as the critical density:

$$\rho_{\text{crit}} \equiv \frac{3H^2 M_{\text{Pl}}^2}{8\pi}. \tag{2.49}$$

In a universe with positive curvature ($\kappa > 0$), the energy density will be greater than the critical density, while a negatively curved universe ($\kappa < 0$) will contain less than this density. It is common for a component of the energy density to be written in terms of its fraction of

the critical density,

$$\Omega_i \equiv \frac{\rho_i}{\rho_{\rm crit}}, \tag{2.50}$$

with $\Omega_{\rm tot} = \rho_{\rm tot}/\rho_{\rm crit}$, where $\rho_{\rm tot}$ is the total density in all forms of energy. Alternatively, this quantity can be related to the curvature and the rate of expansion:

$$\Omega_{\rm tot} = 1 + \frac{\kappa}{a^2 H^2}. \tag{2.51}$$

2.7.1 The Einstein-de Sitter universe

As a first example, let's consider a universe that contains only matter, and that has no curvature. In order for this to be realized, the energy density of the matter must be equal to the critical density, $\rho_M = \rho_{\rm crit}$. In this case, known as the Einstein-de Sitter universe, we can write the energy density as $\rho = \rho_0 a^{-3}$, where ρ_0 is the density of matter evaluated at $a = 1$. Substituting this, along with $\kappa = 0$, into Eq. 2.48, one finds that the scale factor evolves as $a \propto t^{2/3}$:

$$a = \left(\frac{6\pi\rho_0}{M_{\rm Pl}^2}\right)^{1/3} t^{2/3}. \tag{2.52}$$

Looking backward, the scale factor of the Einstein-de Sitter universe was zero a finite time ago (at $t = 0$). By evaluating t at $a = 1$, we can express the age of the universe in terms of the current density or Hubble rate, $t_{\rm age} = M_{\rm Pl}/(6\pi\rho_0)^{1/2} = 2/(3H_0)$. Also notice that the universe expands forever without limit in this case, but with a monotonically decreasing value of the Hubble rate:

$$H \equiv \frac{\dot{a}}{a} = \frac{2}{3t}. \tag{2.53}$$

2.7.2 Models with curvature

Next, let's generalize the Einstein-de Sitter model to allow for the possibility of nonzero curvature. Since κ is no longer zero in this case, it is not trivial to find an analytic solution to Eq. 2.48. Since the first term on the right-hand side evolves as $\rho \propto a^{-3}$ and the second term as a^{-2}, it is clear that the former will dominate if a is sufficiently small, and that the latter will overtake the first term when (or if) the scale factor gets large enough. In the $\kappa < 0$ case (corresponding to $\Omega_M < 1$), the evolution of the universe's scale factor starts out just like one without curvature, $a \propto t^{2/3}$. At later times, however, the curvature term becomes dominant, and a grows in proportion to time:

$$a = \begin{cases} \left(\dfrac{6\pi\rho_0}{M_{\rm Pl}^2}\right)^{1/3} t^{2/3}, & \dfrac{\rho}{M_{\rm Pl}^2} \gg \dfrac{|\kappa|}{a^2} \\ |\kappa|^{1/2}\, t, & \dfrac{\rho}{M_{\rm Pl}^2} \ll \dfrac{|\kappa|}{a^2}. \end{cases} \tag{2.54}$$

This case of $\Omega_M < 1$ (and $\kappa < 0$) is often referred to as an open universe. An open universe expands more rapidly at late times than is predicted in the Einstein-de Sitter model, although still slowing as $H = 1/t$ in the large t limit.

In the $\kappa > 0$ case, the universe does not expand without limit. Instead, space expands until the two terms on the right-hand side of Eq. 2.48 become equal in magnitude, at which point $\dot{a}$ transitions from positive to negative and the universe begins to contract. Ultimately,

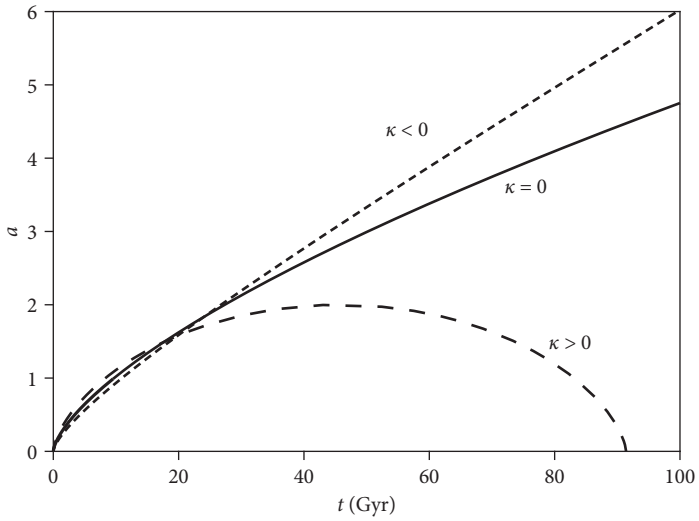


Figure 2.1 The evolution of the scale factor in a matter-dominated universe with $\kappa < 0$, $\kappa = 0$, or $\kappa > 0$. In each case, $H = 67.4$ km/s/Mpc at $a = 1$.

the scale factor becomes zero in this case, returning the universe to a state of infinite density in what is known as a "Big Crunch" (see Figure 2.1).

2.7.3 The de Sitter universe

As another example, we will consider a more exotic universe, which contains no matter or radiation, but whose energy density is instead dominated by a cosmological constant with $\rho_\Lambda = \rho_{\text{crit}}$. In this case, known as the de Sitter universe, space expands exponentially:

$$a = a_0 e^{t/\tau}, \tag{2.55}$$

where a_0 is the scale factor at $t = 0$ and $\tau = (3M_{\text{Pl}}^2/8\pi\rho_\Lambda)^{1/2} = (3/\Lambda)^{1/2}$. Unlike the other cases we have considered, the Hubble rate remains constant in a de Sitter universe, $H = (\Lambda/3)^{1/2}$. Furthermore, while the de Sitter universe is always expanding, it is in a sense not evolving at all. After all, its energy density, composition, and expansion rate all remain unchanged with time.

Although the de Sitter solution might at first glance seem unrelated to our own universe, it will play an important role in our discussions of both inflation and dark energy.

2.7.4 Our universe

None of the idealized scenarios described above can accurately describe the universe we actually live in. Instead of only one substance, our universe currently contains cosmologically relevant quantities of both matter and dark energy, the latter of which can be described in terms of a cosmological constant. Furthermore, our universe was dominated by radiation during the first tens of thousands of years that followed the Big Bang. We can describe the evolving energy density of our universe as follows:

$$\begin{aligned} \rho &= \rho_M + \rho_{\text{rad}} + \rho_\Lambda \\ &= \rho_{\text{crit},0}\,[\Omega_{M,0}\,a^{-3} + \Omega_{\text{rad},0}\,a^{-4} + \Omega_{\Lambda,0}], \end{aligned} \tag{2.56}$$

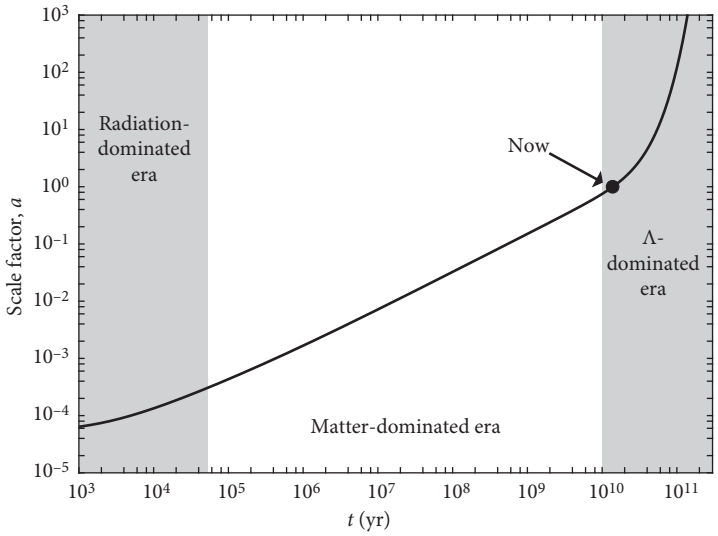


Figure 2.2 The past and future evolution of the scale factor in a universe with our current composition ($\Omega_M = 0.315$, $\Omega_{\rm rad} = 9.26 \times 10^{-5}$, $\Omega_\Lambda = 0.685$) and expansion rate ($H_0 = 67.4$ km/s/Mpc).

where the symbols with a subscript '0' denote the present values of these quantities: $\rho_{\rm crit,0} = 4.94 \times 10^{-6}\,{\rm GeV/cm^3}$, $\Omega_{M,0} = 0.315$, $\Omega_{\rm rad,0} = 9.26 \times 10^{-5}$, and $\Omega_{\Lambda,0} = 0.685$. Dark energy has dominated the energy density of our universe since $a \approx (\Omega_{M,0}/\,\Omega_{\Lambda,0})^{1/3} \approx 0.77$, corresponding to 3.46 billion years ago, or 10.32 billion years after the Big Bang. Looking back further, radiation in the form of photons and neutrinos made up the dominant contribution to our universe's energy density prior to $a \approx \Omega_{\rm rad,0}/\,\Omega_{M,0} \approx 2.94 \times 10^{-4}$, corresponding to 51,100 years after the Big Bang. This transition took place at what is known as the time of matter-radiation equality (see Figure 2.2).

During the radiation-dominated era, the energy density evolved as $\rho \propto a^{-4}$, leading to $a \propto t^{1/2}$. The Hubble rate during this period was related to the age of the universe by $H = 1/(2t)$. As our universe has expanded over cosmic history, not only has its composition changed, but also its temperature. We'll explore the thermal history of our universe in some detail in later chapters of this book.

2.8 Distances in an expanding universe

Before moving on, it will be helpful to briefly consider how cosmologists approach the idea of distance. Distances in cosmology are complicated by the fact that our universe is expanding. For example, if someone were to tell you that a particular galaxy is 5 Gpc away, it would not necessarily be clear what they mean. They might be trying to say that the distance separating us from that galaxy *right now* is 5 Gpc, in which case they are referring to the concept of proper distance. But this is not the only way that this statement could be reasonably interpreted. After all, when the light that we are observing now was emitted from that galaxy, the scale factor was significantly less than 1, and thus the photons have actually moved along a path that is significantly shorter than the proper distance. For these and other reasons, cosmologists employ several different definitions of distance, depending on the application at hand.

As a first case, we can define the comoving distance to an object located at a redshift, z, as follows:

$$\begin{aligned} d_C(z) &= \int_{t(a)}^{t_0} \frac{c\,dt'}{a(t')} \\ &= \int_a^1 \frac{c\,da'}{a'^2\,H(a')}, \end{aligned} \tag{2.57}$$

where $a = (1+z)^{-1}$ and H is the Hubble rate.

The comoving distance is closely related to the concept of the comoving horizon (also known as the particle horizon, or cosmological horizon). To understand this idea, consider how far a photon (or other relativistic particle) could have traveled since the Big Bang if it had undergone no interactions:

$$\begin{aligned} d_h &= \int_0^t \frac{c\,dt'}{a(t')} \\ &= \int_0^1 \frac{c\,da}{a^2\,H(a)} \\ &= \frac{c}{H_0}\int_0^1 \frac{da}{[\Omega_M a + \Omega_{\rm rad} + \Omega_\Lambda a^4]^{1/2}} \\ &\approx 13.4\,{\rm Gpc}, \end{aligned} \tag{2.58}$$

where in the last line I have used measured values of Ω_M, $\Omega_{\rm rad}$, Ω_Λ, and H_0.[2] We are surrounded by a sphere of radius d_h, beyond which no energy or information could have possibly reached us. From any space beyond this horizon, we would seem to be entirely causally disconnected.

Another common way of defining distances in cosmology is in terms of an object's angular size. In particular, we define the angular diameter distance to an object as follows:

$$d_A(z) \equiv \frac{l}{\theta} = \begin{cases} \dfrac{c}{(1+z)H_0\sqrt{|\Omega_\kappa|}}\sin\left(\sqrt{|\Omega_\kappa|}d_C H_0/c\right), & \kappa > 0 \\ \dfrac{d_C}{1+z}, & \kappa = 0 \\ \dfrac{c}{(1+z)H_0\sqrt{\Omega_\kappa}}\sinh\left(\sqrt{\Omega_\kappa}d_C H_0/c\right), & \kappa < 0 \end{cases} \tag{2.59}$$

where l is the physical size and θ is the angle subtended by the object in question. The degree of curvature is quantified in terms of $\Omega_\kappa \equiv 1 - \Omega_0$, where Ω_0 is the value of $\Omega_{\rm tot}$ evaluated today. From this definition of Ω_κ, it follows that $\kappa > 0$ if $\Omega_\kappa < 0$ and $\kappa < 0$ if $\Omega_\kappa > 0$.

One thing to notice is that for relatively small distances, $d_C \ll c/(H_0\sqrt{|\Omega_\kappa|})$, all three of these expressions reduce to $d_A \approx d_C/(1+z)$, making the effects of curvature apparent only on large distance scales. That being said, the angular size of a very distant object will

[2] It might seem surprising that light could have reached us from such a distance, given that 13.4 Gpc ($\approx$ 44 billion light years) is much farther than light can travel in a time equal to the age of the universe. This apparent contradiction is made possible by the fact that our universe was smaller in the past than it is today.

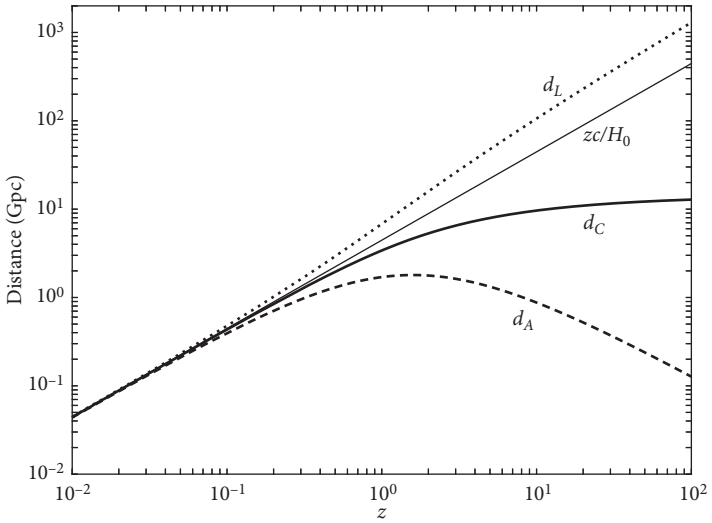


Figure 2.3 The comoving distance, angular diameter distance, and luminosity distance to an object at redshift z (for $\Omega_M = 0.315$, $\Omega_{\rm rad} = 9.26 \times 10^{-5}$, $\Omega_\Lambda = 0.685$, and $H_0 = 67.4$ km/s/Mpc).

depend on the large-scale geometry of our universe. For example, the angular size of the most prominent features in the cosmic microwave background (corresponding to the first peak in its angular power spectrum) is about ~ 1 degree. Since we can calculate the physical size of these regions, this angular scale provides us with a direct measurement of Ω_κ (see Chapter 7). In fact, modern measurements allow us to constrain $-0.003 < \Omega_\kappa < 0.005$ [8]. On cosmological scales, our universe is very flat, showing no signs of any spatial curvature.

A third way of defining the distance to an object is in terms of its perceived brightness. In particular, the observed flux from an isotropically emitting source, F, is related to its intrinsic luminosity, L, and its distance in a way that we can use to define the luminosity distance to the object:

$$d_L \equiv \left(\frac{L}{4\pi F}\right)^{1/2}. \tag{2.60}$$

The energy per unit time that passes through a given spherical shell around an object is reduced as a result of the expansion of space for two different reasons. First, by increasing the scale factor, the number of photons that pass through the shell per unit time is reduced. Second, the energy of each photon is reduced due to cosmological redshift, further decreasing the total power that passes through the shell. Taking these factors into account, we find the following expression for the luminosity distance:

$$d_L(z) = \begin{cases} \dfrac{c\,(1+z)}{H_0\sqrt{|\Omega_\kappa|}}\sin(\sqrt{|\Omega_\kappa|}\,d_C H_0/c), & \kappa > 0 \\ d_C\,(1+z), & \kappa = 0 \\ \dfrac{c\,(1+z)}{H_0\sqrt{\Omega_\kappa}}\sinh(\sqrt{\Omega_\kappa}\,d_C H_0/c), & \kappa < 0. \end{cases} \tag{2.61}$$

Note that, regardless of curvature, the angular diameter and luminosity distances are related as $d_A = d_L/(1+z)^2$.

Additional reading

There are many great textbooks on the subject of cosmology. I'm particularly fond of Dodelson and Schmidt's *Modern Cosmology* and Baumann's *Cosmology*. I also find myself frequently referencing Weinberg's *Cosmology*. For an approachable introduction to general relativity, I recommend both Carroll's *Spacetime and Geometry* and Schutz's *A First Course in General Relativity*. For a more advanced treatment, Wald's *General Relativity* is a popular resource. For a great introduction to cosmology at the undergraduate level, I recommend Ryden's *Introduction to Cosmology*.

Problems

2.1 We will use the system of natural units throughout much of this book. In this spirit, calculate the mass of the Sears Tower in GeV, the distance to Chicago (from wherever you are) in GeV^{-1}, how long it would take you to get there by plane in GeV^{-1}, and Chicago's population per surface area in GeV^2.

2.2 Consider an infinitely large, static, Euclidean universe, containing an average of one galaxy per Mpc^3, each with a luminosity of $L \sim 10^{11}\, L_\odot$, where $L_\odot$ is the luminosity of the sun. If the universe is 10 Gyr old, what is the total flux received on earth from all external galaxies throughout the universe? What if the universe is 100 Gyr old? What if it is infinitely old?

2.3 Perform a change of coordinates to show that the metric given in Eq. 2.6 is homogeneous and isotropic.

2.4 Consider the following expressions involving tensors, vectors, and one-forms: $A^\mu + B_\mu, A_{\mu\nu} = B_{\mu\nu}$, $A_{\mu\nu}B^{\mu\nu}C_{\mu\nu}$, $A^\mu_\nu B^\nu + C^\mu$, and $A_{\mu\nu} = B_\alpha$. Which of these don't make any sense?

2.5 What is the value of δ^μ_μ?

2.6 Show that the following expressions are equivalent to Maxwell's equations:

$$\frac{\partial F_{\lambda\mu}}{\partial x^\nu} + \frac{\partial F_{\mu\nu}}{\partial x^\lambda} + \frac{\partial F_{\nu\lambda}}{\partial x^\mu} = 0 \quad \text{and} \quad \frac{\partial F^{\mu\nu}}{\partial x^\nu} = 4\pi J^\mu,$$

where the Maxwell field strength tensor is related to the electric and magnetic fields as

$$F_{\mu\nu} = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & -B_z & B_y \\ -E_y & B_z & 0 & -B_x \\ -E_z & -B_y & B_x & 0 \end{pmatrix},$$

and $J^\mu = (\rho, \vec{J})$, where ρ and $\vec{J}$ are the charge density and current, respectively.

2.7 From the Friedmann-Leîmatre-Robertson-Walker metric and the expression for the Christoffel symbols given in Eq. 2.17, derive their components as given in Eq. 2.19.

2.8 Derive the R_{11}, R_{22}, and R_{33} components of the Ricci tensor.

2.9 Consider a collection of particles, each with the same amount of energy and mass. What is their equation of state?

2.10 If Einstein's static cosmological solution had described our universe, we could observe our own galaxy from the outside by looking "around" the universe. How difficult would this be? What kind of telescope would be needed to accomplish this task?

2.11 From the Einstein equations including the cosmological term (Eq. 2.38), derive the Friedmann equations (Eqs. 2.39 and 2.40). Hint: Begin by incorporating Λ into the energy-momentum tensor, with $\rho \to \rho + \Lambda M_{\rm Pl}^2/8\pi$ and $p \to p - \Lambda M_{\rm Pl}^2/8\pi$.

2.12 Using Eqs. 2.39 and 2.40, write a program to calculate how a universe with positive spatial curvature, a cosmological constant, and matter evolves. Set the initial matter density to $\rho = (\Lambda + \epsilon)M_{\rm Pl}^2/4\pi$, where $|\epsilon| \ll \Lambda$. If ϵ is exactly zero, you should find the universe to be static. What happens if $\epsilon = 0.01\Lambda$? Or if $\epsilon = -0.01\Lambda$? For a value of $\epsilon/\Lambda = 0.1$, how long would it take for the scale factor to change by a factor of 2? Use the modern measured value for the matter density.

2.13 For the case of a matter-dominated universe with no curvature, $\Omega_M = \Omega_{\rm tot} = 1$, use the first Friedmann equation to show that the age of the universe is related to the current Hubble rate by $t_{\rm age} = 2/(3H_0)$.

2.14 Write a program to calculate the relationship between time and scale factor for any combination of the current densities of matter, radiation, and the cosmological constant (and thus for any curvature). Use it to reproduce the results shown in Figure 2.2.

2.15 Challenge problem: Assuming matter domination, there exist analytic solutions to the first Friedmann equation, even if $\kappa \neq 0$. For $\kappa < 0$, the scale factor evolves as follows:

$$a = \frac{4\pi\rho_0}{3M_{\rm Pl}^2}(\cosh\eta - 1),$$

where ρ_0 is the density of matter at $a = 1$, and η is a parameter related to time as follows:

$$t = \frac{4\pi\rho_0}{3M_{\rm Pl}^2}(\sinh\eta - \eta).$$

Similarly, in the $\kappa > 0$ case,

$$a = \frac{4\pi\rho_0}{3M_{\rm Pl}^2}(1 - \cos\eta),$$

where η is defined in this case as

$$t = \frac{4\pi\rho_0}{3M_{\rm Pl}^2}(\eta - \sin\eta).$$

Show that these solutions follow from the first Friedmann equation. Hint: Solve the first Friedmann equation for time in terms of an integral over the scale factor, and then make the substitution $a(\eta) = (8\pi\rho_0/3M_{\rm Pl}^2\kappa)\sin^2(\sqrt{\kappa}\eta/2)$.

2.16 Calculate the angular diameter distance to a surface at $z = 1090$ for a universe that contains only matter with $\Omega_M = 0.5$, 1, or 1.5. For each value of Ω_M, determine the angular size of a region at $z = 1090$ with a radius equal to the Hubble length, $l_H(z) = c/H(z)$.

3

A Crash Course in Particle Physics

This chapter is intended primarily for readers who do not already have a strong background in particle physics or quantum field theory. While not a substitute for a good textbook on these subjects, this chapter should provide an introduction that will serve those readers well throughout the remainder of this book, which includes numerous references to the ideas and concepts discussed here.

3.1 The Standard Model of particle physics

Particle physics is the study of how elementary particles interact with each other, but it is also the study of the different forms that energy can take in our universe. Every particle—whether an electron, photon, quark, or Higgs boson—is simply a localized excitation of a quantum field that extends across all space and time. By applying quantum field theory, we can hope to understand the various kinds of particles that exist in our universe, and use deep principles of symmetry and mathematical self-consistency to restrict and relate the properties of those particles and their interactions.

Our current best understanding of the particles that exist in our universe is codified in the Standard Model of particle physics. In contrast to a theory such as general relativity, the Standard Model was not discovered by any particular individual, and it was not presented in a finalized form all at one time. Instead, the various components of this theory were constructed and assembled by numerous physicists, working over a span of several decades.

The particles described by the Standard Model consist of 12 kinds of fermions and five kinds of bosons. The three forces that are accounted for in the Standard Model (which does not include a description of gravity) are manifested in nature through the exchange of bosons. The electromagnetic force, for example, is the consequence of photons being exchanged between electrically charged particles. The strong force is realized by the exchange of gluons, and the weak force by the exchange of $W^{\pm}$ and Z bosons.[1]

[1] When counting particle types, note that I am not treating particles and antiparticles separately. More generally speaking, I'm counting any collection of states that differ only in their charges as being of the same particle type. If I had counted all such charge assignments as different particles, the Standard Model would contain 12 leptons (rather than six), 36 quarks (rather than six), eight gluons, one photon, two $W^{\pm}$ bosons, one Z boson, and one Higgs boson, for a total of 61 different kinds of particles.

Six of the fermions are quarks—particles that carry both electric charge and the charge associated with the strong force, known as color. As a result of these charges, the quarks experience both the electromagnetic and the strong forces. The other six fermions are the leptons. Half of the leptons—the electron, muon and tau—carry electric charge, while the other three—the neutrinos—are electrically neutral. These 12 fermions are further organized into three families or generations, each of which contains one charged lepton, one neutrino, one quark with a charge of $+2/3$, and one quark with a charge of $-1/3$. These and other properties of the particles contained within the Standard Model are summarized in the tables provided on page xxi.

The weak force acts on all of the quarks and leptons of the Standard Model, so long as their helicity is left-handed (i.e., so long as they are spinning counterclockwise with respect to their direction of motion). Similarly, the weak force only acts on antiquarks and antileptons that are right-handed. The fact that the weak force distinguishes between particles on the basis of the orientation of their angular momentum is one of the more surprising features of the Standard Model. Since right-handed neutrinos and left-handed antineutrinos do not experience any of the three forces of the Standard Model, there is no way to produce or observe such particles in any plausible experiment. In a sense, at least as far as the Standard Model is concerned, right-handed neutrinos and left-handed antineutrinos simply do not exist.

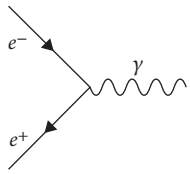
The Standard Model is not a particularly simple or elegant theory, but it is among the most empirically successful in the history of science. It has been used to make numerous highly specific and precise predictions for the ways in which various particles will behave at accelerators and in other laboratory experiments, and so far this theory has passed all such tests. That being said, the Standard Model offers no plausible explanation for the phenomena of dark matter, neutrino masses, the matter-antimatter asymmetry, inflation, or gravity, and thus it cannot possibly be considered complete. Furthermore, the Standard Model has 19 free parameters that are not predicted by the theory and that can only be determined through measurement. These numbers set the masses of various particles and the strengths of the three forces, and parameterize how the different quarks interact through the weak force. Ideally, we would like to understand where these numbers come from and what determines their values, but at this time we can do little more than speculate on this topic.

Although the Standard Model is a complex theory, its structure is far from arbitrary. In fact, the form it takes is highly constrained by the requirements of mathematical self-consistency. In order for a quantum field theory to avoid making infinite predictions for necessarily finite physical quantities, it must be renormalizable, and this has motivated particle physicists to focus on theories that are based on the principle of gauge symmetry. In the case of the Standard Model, the mathematical structure of this theory respects a gauge symmetry that can be written as $SU(3)_C \times SU(2)_L \times U(1)_Y$. While I will not discuss the full meaning of these symbols here, it suffices to say that the requirement of gauge symmetry strongly restricts the forms that a viable particle physics theory might take—including that of the Standard Model.

3.2 Feynman diagrams

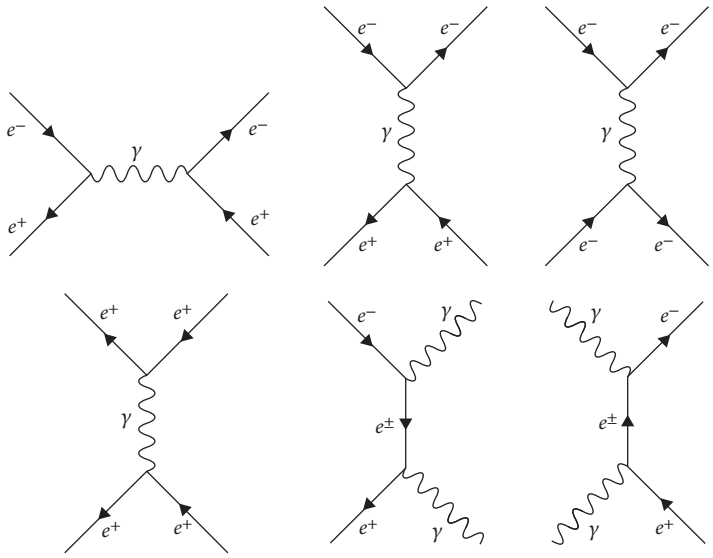
Let's begin this section by considering a simplified version of the theory of quantum electrodynamics (QED), which describes the interactions between photons, electrons, and

positrons. Any interaction that can occur in this theory can be built from combinations of the following vertex:



You can think of this simple vertex as the building block that we will use in this simplified theory to construct all of the interactions that can take place between particles. In the symbolic language of Feynman diagrams, any solid line with an arrow pointing toward the right (taken to be the direction of time) represents a fermion (an electron, in this case), while any solid line pointing to the left represents an antifermion (in this case a positron). It doesn't matter if the lines are also pointing upward or downward. The squiggly line in the vertex shown above represents a photon. Other spin-1 particles such as the $W^{\pm}$ and Z bosons are also represented by this line type. For some reason, spin-1 gluons are represented by a different line type that we will see later. We can also rotate this vertex, allowing us to describe interactions featuring different combinations of particles and antiparticles. For example, by rotating clockwise by 90°, we find ourselves with a vertex connecting an incoming and outgoing positron to a photon. After an additional rotation of 180°, we have a photon coupling to an incoming and outgoing electron.

From this single vertex, we can build any number of possible interactions. By combining two such vertices, for example, we can construct any of the following Feynman diagrams:



The upper left and upper middle diagrams collectively represent the process of electron-positron elastic scattering, $e^+e^- \rightarrow e^+e^-$. The following two diagrams also represent elastic scattering processes, but for a pair of electrons (upper right) or positrons (lower left). The remaining two diagrams correspond to the processes of electron-positron

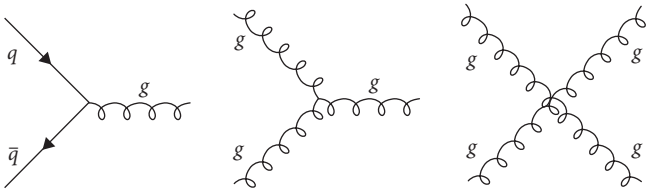
pair annihilation, $e^+e^- \rightarrow \gamma\gamma$, and pair production, $\gamma\gamma \rightarrow e^+e^-$. In addition to those shown above, there are four other two-vertex diagrams, which correspond to the elastic scattering of an electron or a positron with a photon, $e^{\pm}\gamma \rightarrow e^{\pm}\gamma$.

Feynman diagrams with the topology of that shown in the upper left example are called *s*-channel diagrams, which feature a propagator (in this case, the photon) that is oriented in the direction of time. The other five cases shown above are examples of *t*-channel diagrams. For every *t*-channel diagram with identical outgoing particles, there is also a corresponding *u*-channel diagram in which the outgoing lines on the right-hand side cross each other. Note that the existence of a *u*-channel diagram is often taken to be implied, and I will not explicitly show any *u*-channel diagrams in this book. Also note that some processes are described by both *s*-channel and *t*/*u*-channel diagrams (such as $e^+e^- \rightarrow e^+e^-$), while others are described by only one or the other.

Before moving on, a few words about the meaning of Feynman diagrams are in order. The lines that appear in these diagrams do *not* represent the physical trajectories of particles through space. Two particle lines that move away from each other following an interaction do not imply the act of a repulsive force, for example. Instead of thinking in terms of motion through physical space, you should think of Feynman diagrams as symbolic tools for representing the various interactions that can take place between incoming and outgoing particles. In this sense, Feynman diagrams are strictly topological in nature, and depend only on the vertices involved, and on whether any of the lines cross. These diagrams can be used to determine the kinds of interactions that are possible in a given theory, and to calculate the amplitudes for the corresponding processes. These amplitudes can then be used to determine things like scattering cross sections and decay rates.

To move beyond our simplified version of QED and toward a more complete description of the Standard Model, we have to include a variety of other vertices, no longer limiting ourselves to the interactions of electrons and positrons with photons. The photon has an analogous vertex with every electrically charged particle. In other words, the photon couples to all electrically charged particles, or equivalently, the electromagnetic force acts on all particles with electric charge.

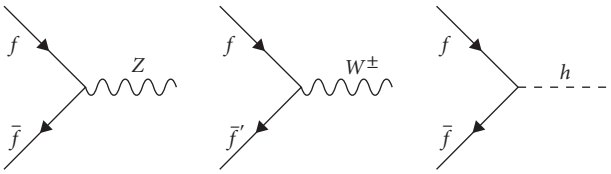
Whereas the electromagnetic force is communicated by the photon, the strong force is manifest through the exchange of gluons. The vertices that combine to build the interactions of quantum chromodynamics (QCD) are as follows:



A key difference between QED and QCD is that whereas there is only one kind of electric charge, there are three kinds of color charge, known as red, green, and blue. Each quark, *q*, carries one unit of one of these three colors (either red, green, or blue), while each antiquark carries a value of −1 of one of these colors (antired, antigreen, or antiblue). If all three colors are combined in equal amounts, they add up to a total of zero net color charge, in analogy to how a mixture of all visible colors produces an image that is white. Each gluon carries one unit of color and one unit of a different anticolor, such as red-antiblue, or green-antired. The fact that the gluon carries color charge further distinguishes it from the photon, which is

electrically neutral. This feature allows the gluon to directly couple to other gluons, whereas photons do not (directly) couple to other photons.

Next, the effects of the weak force are built up out of vertices that include the following:



The $W^{\pm}$ and Z are known as the electroweak gauge bosons, and they couple with similar strength to all of the left-handed Standard Model fermions, f, and right-handed antifermions, $\bar{f}$. While the Z is electrically neutral, the $W^{\pm}$ carries one unit of electric charge. The Higgs boson, h, couples to Standard Model fermions with a strength that is proportional to the fermion's mass. In fact, it is this interaction that causes Standard Model fermions to have a mass. The strength of this coupling, called a Yukawa coupling, determines the masses of the charged leptons and quarks.[2] In addition to the three vertices shown above, there are several others that involve combinations of electroweak gauge bosons and Higgs bosons.

One important distinction between the weak force and the strong or electromagnetic forces is that the *type* of a fermion can be changed at a vertex involving a $W^{\pm}$ boson. In fact, since the $W^{\pm}$ carries electric charge, the electric charges of the f and f' in the $W^{\pm}$ vertex shown above must differ by 1. For example, a $W^{\pm}$ could transform an electron into a neutrino, or an up quark into a down quark. Furthermore, these interactions with a $W^{\pm}$ can also transform a quark of one generation into one of another. The coupling of the $W^{\pm}$ to any two kinds of quarks (or antiquarks) is proportional to the elements of what is known as the Cabibbo-Kobayashi-Maskawa (CKM) matrix, which is given by

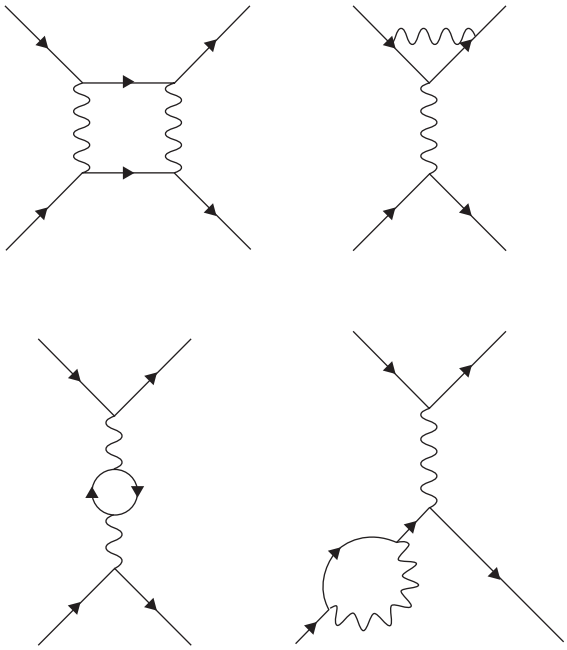
$$\begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} = \begin{pmatrix} 0.974 & 0.227 & 0.004 \\ 0.227 & 0.973 & 0.042 \\ 0.008 & 0.042 & 0.999 \end{pmatrix}. \tag{3.1}$$

The coupling between a $W^{\pm}$ and quarks of type i and j is proportional to V_{ij}, where $i = u$, c, or t and $j = d$, s, or b. The numerical values of the elements of this matrix are measured quantities, and are not predicted by the structure of the Standard Model. Notice that the CKM matrix is roughly diagonal, which means that the two quarks involved in these interactions tend to be of the same generation. That being said, the ability of the weak force to change quarks of one generation into those of another plays an important role in a wide range of physical phenomena.

So far, we have been focused on Feynman diagrams with two vertices, known as tree-level diagrams. In principle, however, a given rate or cross section is impacted by *all* possible Feynman diagrams for that process, including those with more than two vertices. Consider, for example, the elastic scattering of two electrons within the context of our simplified

[2] Strictly speaking, neutrinos should not have any mass according to the Standard Model. The small masses that neutrinos are observed to possess must be generated by some means beyond the Standard Model, and cannot be the result of a simple coupling to the Higgs boson.

version of QED, $e^- e^- \to e^- e^-$. In addition to the two (t- and u-channel) tree-level diagrams for this process described above, we can write down a fairly long list of diagrams with four vertices, a few of which are:



Here, for simplicity, I have not explicitly labeled the lines as electrons or photons. Diagrams with four vertices are called one-loop diagrams, and in this case there are 18 of these diagrams that contribute to the process at hand. In addition, I could write down an even larger number of diagrams with two loops, or three, and so on. It goes without saying that including such a large number of diagrams could quickly make any calculation intractable. In practice, we often limit our calculations to the tree-level contribution, and include those from one or two loop diagrams only in instances that require particularly high levels of precision. The reason that we can, and often do, neglect the higher order diagrams is that in calculating the amplitude for a given process, each vertex carries with it a factor that is proportional to the coupling involved, and in many cases this factor is much less than 1, making the contributions from higher order diagrams small compared to those from lower order diagrams. In this respect, Feynman diagrams represent a useful perturbative expansion. In cases in which this expansion is *not* perturbative (such as when dealing with a theory with a large coupling strength), one cannot rely on the techniques of Feynman calculus, but must take other approaches in order to obtain reliable results.

3.3 Conservation laws

Built into the rules for constructing Feynman diagrams in the Standard Model are a variety of conservation laws. These conservation laws make it easy to identify which processes are and are not possible, such as which particles can decay and which are stable.

In addition to those pertaining to energy and momentum, the following conservation laws are manifest in the Standard Model:

- *Electric charge* is conserved in every interaction in the Standard Model. In the language of Feynman diagrams, the total amount of electric charge that goes into any given vertex is always exactly equal to the charge that goes out. This fact is guaranteed by the gauge symmetry of the theory.
- *Color* is also conserved in every interaction in the Standard Model, being a consequence of the gauge symmetry of $SU(3)_C$. As quarks carry one unit of color (red, green or blue) and gluons carry one unit of color and one of a different anticolor (such as red-antiblue, or green-antired), each *qqg* vertex changes the color of the incoming and outgoing quarks.
- *Baryon number* is a quantum number that is equal to $1/3$ for every quark, $-1/3$ for every antiquark, and 0 for all other particles in the Standard Model. Since every interaction with a quark going into a vertex also has one going out of that vertex, baryon number is conserved in all Standard Model Feynman diagrams. There are, however, non-perturbative processes in the Standard Model that can violate the conservation of baryon number, as I will discuss below.
- *Lepton number* is a quantity similar to baryon number, and is equal to 1 for all charged leptons and neutrinos, and to -1 for their antiparticles. Also like baryon number, lepton number is conserved in every vertex in the Standard Model, but is not conserved in certain non-perturbative processes.
- *Flavor* is a collection of quantum numbers that are associated with the species of a given quark or lepton. Neutrinos, for example, come with subscripts that denote their electron, muon, or tau flavor: ν_e, ν_μ, and ν_τ. All interactions involving the electromagnetic or strong forces conserve flavor in the Standard Model, as these vertices never alter the identity of any fermion. In weak interactions, however, flavor is not necessarily conserved. In particular, the coupling of the $W^\pm$ boson can transform a quark of one generation into a quark of another, according to the CKM matrix.

At temperatures above the electroweak phase transition, $T \gg T_{\rm EWPT} \sim 200\,{\rm GeV}$, non-perturbative processes can take place that violate the conservation of both baryon and lepton number. These processes involve what are known as electroweak sphalerons, which cannot be represented by Feynman diagrams. In such interactions, nine quarks and three leptons are involved in a way that conserves the difference between baryon number and lepton number, $B-L$, but not baryon number or lepton number individually. For example, six up quarks and three down quarks could collectively transform into three positrons, keeping the value of $B-L$ fixed at 3 throughout this interaction (see Sec. 12.4).

It is the conservation of electric charge in the Standard Model that causes the electron (the lightest particle with electric charge) to be absolutely stable. On similar grounds, you might think that the conservation of color would stabilize the lightest colored particle. As I will discuss in the following section, however, all naturally occurring bound states of quarks carry zero total color charge, so this conservation law does not lead to the stability of any particle species.

If baryon number were perfectly conserved, this would stabilize the proton, and it almost does. In principle, protons can be destroyed in the Standard Model through processes involving electroweak sphalerons, but at an *incredibly* suppressed rate, smaller than typical processes mediated by the weak force by a factor of $\sim e^{16\pi^2/g_2^2} \sim 10^{161}$ [14, 15]. The lightest

neutrino, being the lightest particle with either lepton or baryon number, is stable in the Standard Model as a consequence of the conservation of baryon-minus-lepton number. The heavier two neutrinos are presumably unstable, but with extremely slow decay rates, suppressed by factors of $\sim m_W^4 m_\mu / m_\nu^5 \sim 10^{57} \times (0.1\,\text{eV}/m_\nu)^5$ relative to those of typical processes mediated by the weak force [16, 17].

3.4 Phase transitions in the Standard Model

3.4.1 Baryons, mesons, and the QCD phase transition

In describing the particles contained within the Standard Model, I have barely mentioned protons or neutrons, as you may have noticed. This is because protons and neutrons are not elementary particles, but are instead composite states, made up of quarks bound together by the strong force. In this section, I will describe the relationship between quarks and the bound states they can form, known as mesons and baryons.

A peculiar feature of the strong force is that its strength (i.e., the value of its coupling, g_s) is larger at long distances than it is at short distances (or equivalently, it is weaker at high energies). This running of the strong coupling happens for the same reason that the effective electric charge of an object depends on the distance from it—it is screened by the distribution of virtual charged particles that surround it. The analogous screening of color charge also occurs in QCD, but with the important difference that gluons themselves carry color charge, and thus participate in the screening. When this calculation is carried out, one finds that the electromagnetic coupling is modestly stronger at short distances than it is at long distances. In contrast, it turns out that the strong coupling becomes significantly weaker when acting on bodies that are very close to one another.[3] At distances shorter than $\sim 10^{-15}$ m, corresponding to energies above what is known as the QCD scale, $\Lambda_{\text{QCD}} \approx 0.3\,\text{GeV}$, the strong coupling is small enough that we can treat interactions involving the strong force perturbatively, allowing us to think about quarks and gluons as free particles. At greater distances (corresponding to energies below Λ_{QCD}), the strong force is powerful enough to bind quarks and gluons together into mesons and baryons, a phenomena known as color confinement.

Roughly $\sim 10^{-5}$ seconds after the Big Bang, at a temperature of $T \sim \Lambda_{\text{QCD}}$, our universe underwent what is known as the QCD phase transition. Under the conditions that existed prior to this transition ($T \gtrsim \Lambda_{\text{QCD}}$), the theory of QCD describes quarks, gluons, and their interactions. At lower energies or temperatures ($T \lesssim \Lambda_{\text{QCD}}$), the same theory describes a collection of hadrons, which are further classified as mesons (bound quark-antiquark pairs, $q\bar{q}$), baryons (bound states of three quarks, qqq), and antibaryons (bound states of three antiquarks, $\bar{q}\bar{q}\bar{q}$). A peculiar feature of QCD is that all naturally occurring hadrons carry zero net color charge—they must be "colorless." For this reason, the quark and antiquark that make up a given meson must have equal and opposite color charges, such as red-antired, or blue-antiblue. On similar grounds, all baryons must contain one quark of each color (red, green, and blue), for a total of zero net color charge.

3.4.2 The electroweak phase transition

At temperatures below ~ 200 GeV, particles experience the electromagnetic and weak forces as mediated through the exchange of photons and $W^\pm$ and Z bosons, respectively. At higher

[3] A gauge theory that becomes increasingly weak at short distances, such as QCD, exhibits what is known as asymptotic freedom [18, 19].

energies or temperatures, however, the character of these forces is very different. Much as how the theory of QCD describes mesons and baryons at low temperatures, and quarks and gluons at high temperatures, the bosons known as the photon, $W^{\pm}$, and Z only exist at temperatures below the electroweak phase transition. Above this transition, these same degrees of freedom take the form of other particles, which collectively communicate the effects of the electroweak force. The Standard Model is built upon the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$. The $SU(3)_C$ part of this expression corresponds to the structure of QCD, which remains unchanged by the electroweak phase transition. The group $SU(2)_L \times U(1)_Y$ describes two forces that are present in the high-energy limit, both of which are mediated by massless gauge bosons. There is exactly one massless gauge boson associated with $U(1)_Y$, usually denoted by B, which couples to particles with a strength that is proportional to their hypercharge, Y. The group $SU(2)_L$ corresponds to three massless gauge bosons, W^1, W^2, and W^3, each of which couples to weak isospin T. All of the Standard Model fermions carry hypercharge, and all of the left-handed Standard Model fermions carry weak isospin (thus the L subscript in the group, $SU(2)_L$).

In addition to these gauge bosons, the Standard Model contains a Higgs boson and associated Goldstone bosons whose properties are dictated by the symmetry of $SU(2)_L$. Unlike the other fields in the Standard Model, the Higgs field has a nonzero value in its ground state. More specifically, it has a vacuum expectation value (vev) of $v = 246$ GeV. At temperatures well above the Higgs vev, $T \gg v$, all of the particles of the Standard Model are massless, including the leptons, quarks, and gauge bosons. As the temperature drops below this value, however, the electroweak sector of the Standard Model undergoes the process of spontaneous symmetry breaking (see Sec. 12.6), transforming it in many respects. In particular, at low temperatures, $T \lesssim v$, the quarks and charged leptons interact with the Higgs field in a way that causes them to behave as massive particles. In other words, it is as a result of these interactions that these particles obtain their masses, through what is known as the Higgs mechanism. The W^1 and W^2 also obtain a mass, to become the charged $W^{\pm}$ boson. Furthermore, during this transition, the B and W^3 bosons cease to exist as independent entities, but instead combine, or mix, to form the massive Z boson, and the massless photon:

$$\begin{aligned} Z &= W^3 \cos\theta_W - B \sin\theta_W, \\ \gamma &= W^3 \sin\theta_W + B \cos\theta_W, \end{aligned} \tag{3.2}$$

where θ_W is the Weinberg angle (also known as the weak mixing angle), which quantifies the degree of mixing between the electroweak gauge bosons. The value of the Weinberg angle is related as follows to the strengths of the couplings associated with the $U(1)_Y$ and $SU(2)_L$ gauge groups, g_1 and g_2:

$$\sin^2\theta_W = \frac{g_1^2}{g_1^2 + g_2^2} \approx 0.231. \tag{3.3}$$

The interactions of the Z and the photon are thus inherited from those of the W^3 and B, weighted by the factors of $\cos\theta_W$ and $\sin\theta_W$ as described in Eq. 3.2. Whereas the electroweak bosons that exist at high temperatures couple to the quantum numbers of hypercharge and weak isospin, the photon and the Z couple to combinations of these quantities. Electric charge, for example, is related to hypercharge and weak isospin by $Q = T^3 + Y$, where $T^3 = T$ for neutrinos and up-type quarks (those with $Q = +2/3$) and $T^3 = -T$ for charged leptons and down-type quarks.

A number of highly precise and specific predictions can be derived from the structure of the Standard Model's electroweak sector. The masses and coupling strengths of the photon,

Z, and $W^{\pm}$, for example, can each be calculated from the value of the Higgs vev and the gauge couplings, g_1 and g_2 [20]. These predictions have been ruthlessly tested at particle accelerators, confirming the validity of the Standard Model to a remarkably high degree of precision. While the Higgs mechanism is also responsible for the masses of the charged leptons and quarks, the value of each of these masses depends on an independent free parameter (the Yukawa coupling of the Higgs to that fermion), and thus the electroweak theory is not particularly predictive in this respect.

3.5 Amplitudes

The tools of quantum field theory can be used to make powerful predictions for any number of scattering cross sections and decay rates. While many of the details are beyond the scope of this chapter, I'm going to try to convey here some of the main elements that go into the calculations of these quantities.

To determine the decay rate or cross section for a given process, one must first calculate the corresponding amplitude, $\mathcal{M}$, which contains all of the relevant information pertaining to the dynamics at hand. To calculate an amplitude, one uses what are known as the Feynman rules of the theory to determine the contribution from each diagram. This involves writing down factors for each incoming particle, outgoing particle, propagator (any particle that is neither incoming nor outgoing), and vertex. This quantity is then integrated with respect to the momenta of any propagators to obtain the amplitude. Each Feynman diagram for a given process contributes a separate term to the amplitude. Next, the absolute square of the amplitude is calculated, $|\mathcal{M}|^2$, averaging over the spin states of the incoming particles, and summing over the possible outgoing spin states. This is then combined with phase space factors to calculate the rates for any corresponding scattering or decay processes. Note that for any process, the dimensionality of the amplitude is related to the total number of incoming plus outgoing particles, n, according to the following: $[\mathcal{M}] = E^{4-n}$. For example, the dimensions of the amplitude for a two-body decay (with one incoming and two outgoing particles) are those of energy. In contrast, the amplitude for any scattering processes with two incoming and two outgoing particles will be unitless.

This book will not describe how to calculate amplitudes in detail; this information can be found in any number of textbooks on particle physics or quantum field theory. Instead, my intention here is to give the reader some intuition for how these calculations are carried out, and to provide you with enough information to make reasonable estimates for various cross sections and decay rates. To get a rough estimate for the amplitude associated with a given process, one can apply the following procedure.

1. Each vertex of a Feynman diagram contributes a factor to the amplitude that is proportional to the coupling and charges involved. For each vertex with a photon, for example, one includes a factor of $g_2 \sin\theta_W \approx 0.31$, multiplied by the electric charge of the interacting particle. For each vertex involving gluons and quarks, one includes an analogous factor that, up to order 1 factors, is equal to half of the strong coupling, $g_s/2$, where g_s varies from $g_s \sim 3$ at energies near $E \sim \Lambda_{\rm QCD}$, to $g_s \sim 1.2$ around $E \sim 100\,{\rm GeV}$. Vertices connecting a $W^{\pm}$ or Z boson to leptons or quarks contribute factors of approximately $g_2\sqrt{2}/4 \approx 0.23$ and $g_2/4\cos\theta_W \approx 0.19$, respectively. When dealing with calculations involving g_2, it is common to write these factors in terms of the fine structure constant, $\alpha \equiv g_2^2 \sin^2\theta_W/4\pi \approx 1/137$, or the Fermi coupling constant, $G_F \equiv \sqrt{2}g_2^2/8m_W^2 \approx 1.16 \times 10^{-5}\,{\rm GeV}^{-2}$.

2. Each propagator in a Feynman diagram contributes a factor that is proportional to 1 over the square of the four-momentum of the propagator minus the square of its mass, $(p^2 - m^2)^{-1}$. Four-momentum will be further discussed in the following section. In the high-energy ($E \gg m$) limit, this factor reduces to $\sim E^{-2}$, while in the low-energy limit it is given by $\sim m^{-2}$.

3. Many amplitudes are further multiplied by factors of the total energy. Use the fact that the dimensionality of a given amplitude is related to the total number of incoming plus outgoing particles, $[\mathcal{M}] = E^{4-n}$, to determine how many of these factors should be included.

While one should not expect this procedure to yield a particularly accurate expression for a given amplitude, it does generally allow one to obtain a reasonable estimate. As an example, let's use this procedure to estimate the amplitude for Compton scattering, $\gamma e^- \to \gamma e^-$. In this case, the two vertices in the tree-level Feynman diagrams each contribute a factor of $g_2 \sin\theta_W$, while the propagator contributes $E_{\rm tot}^{-2}$. Since this process has a total of four incoming and outgoing particles, its amplitude should be unitless, leading us to include an additional at factor of $E_{\rm tot}^2$, and thus to estimate $|\mathcal{M}|^2 \sim g_2^4 \sin^4\theta_W$. In Sec. 3.7, we'll see how to use amplitudes such as this to calculate scattering cross sections and decay rates for various processes.

In the first step of the procedure described above, you might be surprised to notice that the vertex factors associated with the weak force are similar in magnitude to those arising from the electromagnetic force. If this were the whole story, the weak force would be approximately as strong as that of electromagnetism. The reason that the weak force is well described by its name is that at low energies, corresponding to $p^2 \ll m_{W,Z}^2$, the propagator factor that goes into this calculation is equal to $m_{W,Z}^{-2}$, leading to amplitudes that are suppressed relative to those associated with the electromagnetic force by a factor of $\sim (E/m_{W,Z})^2$. Because of this suppression, the cross section for a low-energy neutrino to scatter with an electron is smaller than that for electron-electron scattering by a factor of roughly $\sim (0.2/0.31)^4 (m_e/m_Z)^4 \sim 10^{-22}$. At energies comparable to or higher than $m_{W,Z}$, however, the weak force is not suppressed by such factors, typically leading to decay rates and scattering cross sections that are similar to those attributable to the electromagnetic force.

Another way to think about the difference between the strengths of the electromagnetic and weak forces is in terms of their range. A consequence of the Heisenberg uncertainty principle, $\Delta E \Delta t \sim \hbar/2$, is that a particle of energy, E, can exist as a virtual particle for a time, $\Delta t \sim \hbar/2E$. During this time, such a particle can travel up to a distance of $d \sim c\Delta t \sim \hbar c/2m$, where m is the mass of the particle in question. Because the photon is massless, electromagnetism is a long-ranged force. The considerable masses of the $W^\pm$ and Z bosons, in contrast, limit the influence of the weak force to a range of $\sim 1/(2m_{W,Z}) \sim 10^{-16}$ cm. From this argument, you might think that the massless nature of the gluon would make the strong force long-ranged, but because they carry color, gluons cannot exist in isolation. As a result, the range of the strong force is limited by the mass of the lightest colorless combination of particles that carry QCD color, the neutral pion, whose mass is $m_{\pi^0} \approx 0.135$ GeV. This leads to a range for the strong force of $\sim 10^{-13}$ cm. Not coincidentally, this is approximately the size of a typical atomic nucleus.

3.6 Relativistic kinematics and four-vectors

When considering the kinematics of relativistic particles, it is often useful to employ what are known as four-vectors. Let's start with the position-time four-vector, which is defined

as $x^\mu \equiv (t, x, y, z)$. We can relate this point in spacetime to the corresponding four-vector in another frame of reference by applying the Lorentz transformation,

$$x^\mu = \Lambda^\mu_\nu x^\nu, \tag{3.4}$$

where I remind the reader that the summation over $\nu = 0, 1, 2, 3$ is implied, as described in Sec. 2.2. The quantity Λ^μ_ν is a 4×4 matrix that represents the Lorentz transformation. For example, for frames of reference moving relative to each other in the x direction:

$$\Lambda^\mu_\nu = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{3.5}$$

Although the entries of a given four-vector are generally different when evaluated in different frames of reference, there are certain combinations of these elements that are always the same, regardless of the frame of reference. Such quantities are called Lorentz invariants. For example, any spacetime interval, $t^2 - x^2 - y^2 - z^2$, is a Lorentz invariant. Using the Minkowski metric (i.e., the metric tensor in flat spacetime with $a = 1$), we can write this invariant compactly as $g_{\mu\nu} x^\mu x^\nu$. Since we can use the metric to lower one of these indices, $x_\mu = g_{\mu\nu} x^\nu$, we can also write this invariant as $x_\mu x^\mu$ (see Sec. 2.2).

The position-time four-vector is merely one example of a four-vector. More generally speaking, any four-component object is a four-vector if it transforms in the same way that x^μ does when moving between different inertial frames of reference (as described by Eq. 3.4). A particularly useful example is the energy-momentum four-vector, $p^\mu \equiv (E, p_x, p_y, p_z) \equiv (E, \vec{p}\,)$, also known as the four-momentum. Just as with the position-time four-vector, the square of the energy-momentum four-vector is a Lorentz invariant:

$$p_\mu p^\mu = E^2 - \vec{p}^{\,2} = m^2, \tag{3.6}$$

where m is the mass of the object described by the four-vector.

To see why the energy-momentum four-vector is so useful, consider the example of pion production in the collision of a cosmic-ray proton with a proton at rest, $p_{\rm CR} + p_{\rm rest} \rightarrow p + p + \pi^0 + \pi^0 + \cdots$. In order for this process to take place, the incoming cosmic ray must have enough energy to create the extra mass of the pions. To determine the value of this threshold energy for pion production, let's look at the energy-momentum four-vector in the rest frame of the target proton:

$$p^\mu_{\rm tot} = (E_{\rm CR} + m_p, |\vec{p}_{\rm CR}|, 0, 0). \tag{3.7}$$

In the center-of-momentum frame, the energy-momentum four-vector evaluated at the threshold energy (at which all of the final state particles are at rest) is given by

$$p^\mu_{\rm tot} = (2m_p + Nm_\pi, 0, 0, 0), \tag{3.8}$$

where N is the number of pions that are produced. Taking advantage of the fact that $p_\mu p^\mu$ is a Lorentz invariant, we can calculate this quantity for both of the above four-vectors and set them equal to each other, resulting in the following:

$$(E_{\rm CR} + m_p)^2 - \vec{p}^{\,2}_{\rm CR} = (2m_p + Nm_\pi)^2 \tag{3.9}$$

$$\longrightarrow E_{\rm CR}^2 + m_p^2 + 2E_{\rm CR}m_p - (E_{\rm CR}^2 - m_p^2) = 4m_p^2 + N^2 m_\pi^2 + 4Nm_p m_\pi$$

$$\longrightarrow E_{\rm CR} = m_p + 2Nm_\pi + \frac{N^2 m_\pi^2}{2m_p}.$$

For example, in order for such a collision to produce a single pion, the incoming cosmic ray must have a kinetic energy in excess of $E_{\rm kin} > 2m_\pi + m_\pi^2/2m_p \approx 0.28\,\text{GeV}$. To produce three pions, $E_{\rm kin} > 6m_\pi + 9m_\pi^2/2m_p \approx 0.90\,\text{GeV}$.

3.7 Decays and scattering

3.7.1 Decay rates

Once the amplitude for a given process has been calculated, we can integrate over the momenta and energies of the incoming and outgoing particles (i.e., over the phase space) to determine the associated decay rate or scattering cross section. The rate for a particle decay, $1 \to 2 + 3 + 4 + \cdots + n$, is related to the corresponding squared amplitude as follows:

$$\Gamma = \frac{S}{2m_1}\int |\mathcal{M}|^2 (2\pi)^4 \delta^4(p_1 - p_2 - p_3 \cdots - p_n) \times \prod_{j=2}^{n} 2\pi\,\delta(p_j^2 - m_j^2)\theta(E_j)\frac{d^4 p_j}{(2\pi)^4}, \quad (3.10)$$

where m_i and $p_i \equiv (E_i, \vec{p}_i)$ are the mass and four-momentum of the ith particle, and S is a statistical factor that is equal to the product of $1/s!$ for each group of s identical outgoing particles. The rightmost delta function enforces $E^2 = \vec{p}^{\,2} + m^2$ for each outgoing particle, while the theta function (a Heaviside step function) ensures that each particle's energy is positive in magnitude. The leftmost delta function ensures that both energy and momentum are conserved. After some manipulation, this equation can be rewritten as

$$\Gamma = \frac{S}{2m_1}\int |\mathcal{M}|^2 (2\pi)^4 \delta^4(p_1 - p_2 - p_3 \cdots - p_n) \times \prod_{j=2}^{n} \frac{1}{2\sqrt{\vec{p}_j^{\,2} + m_j^2}}\frac{d^3\vec{p}_j}{(2\pi)^3}. \quad (3.11)$$

In the case of a two-body decay, $1 \to 2 + 3$, evaluated in the rest frame of the decaying particle, this simplifies to

$$\Gamma = \frac{S\,|\vec{p}\,|}{8\pi m_1^2}|\mathcal{M}|^2, \quad (3.12)$$

where $\vec{p}$ is the momentum of either of the two decay products.

From this expression, we can see how the masses of the particles involved can impact a decay rate. For cases with light decay products, $m_1 \gg m_2, m_3$, energy and momentum conservation require $|\vec{p}\,| \simeq m_1/2$, and thus $\Gamma \propto |\mathcal{M}|^2/m_1$. Furthermore, since $\mathcal{M} \propto m_1$ in most cases, we generally expect $\Gamma \propto m_1$, explaining why heavier particles tend to decay more rapidly than lighter particles. Also notice that decay rates will be suppressed if the sum of the final state particle masses is not much less than the mass of the decaying particle, leading to $|\vec{p}\,| \ll m_1/2$.

Let's put together what we have learned so far to make an order of magnitude estimate for a decay rate. Consider, for example, the rate for the decay of a Z boson to an

electron-positron pair, $Z \to e^+e^-$. This process has one vertex, which contributes a factor of $\sim g_2/4\cos\theta_W$. Since this process has a total of $n = 3$ incoming and outgoing particles, the amplitude must have dimensions of energy, and since the only thing that sets an energy scale in the problem is the mass of the Z, we can infer that the squared amplitude should be roughly $|\mathcal{M}|^2 \sim g_2^2 m_Z^2/16\cos^2\theta_W$. Combining this with Eq. 3.12 and taking $|\vec{p}\,| \approx m_Z/2$, we estimate $\Gamma(Z \to e^+e^-) \sim m_Z g_2^2/256\pi\,\cos^2\theta_W$, which is not too far from the correct answer of $\Gamma(Z \to e^+e^-) \approx (m_Z g_2^2/192\pi\,\cos^2\theta_W) \times [1 + (4\sin^2\theta_W - 1)^2]$. Note that we could perform a similar estimate for the decay rate of a Z boson to a quark-antiquark pair, or a neutrino-antineutrino pair. In the case of the decay into quarks, we must include an additional factor of 3 in order to account for the production of quarks of each color.

Before moving on to scattering cross sections, I want to introduce a few definitions. First, the lifetime of a particle is defined as the inverse of its total decay rate, $\tau = \Gamma^{-1}$. Second, the total decay rate of a particle is often referred to as its decay width. The width of a particle can be thought of as a measure of the quantum uncertainty associated with its mass, following from the Heisenberg uncertainty principle, $\Delta E \Delta t \sim \hbar/2$. Next, the rate for a particle to decay into a particular final state is called a partial width, and the ratio of a partial width to the total width (i.e., the probability that the decay will proceed to that final state) is known as a branching fraction.

3.7.2 Scattering cross sections

Next, consider the case of two particles scattering, $1 + 2 \to 3 + 4 + \cdots + n$. The cross section for this process is related to the corresponding amplitude as follows:

$$\sigma = \frac{S}{4\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2)^2}} \int |\mathcal{M}|^2 (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4 \ldots p_n) \times \prod_{j=3}^{n} 2\pi\,\delta(p_j^2 - m_j^2)\theta(E_j)\frac{d^4\vec{p}_j}{(2\pi)^4}, \tag{3.13}$$

which can be rewritten in terms of an integral over the outgoing particles' momenta, $\vec{p}_j$:

$$\sigma = \frac{S}{4\sqrt{(p_1 \cdot p_2)^2 - (m_1 m_2)^2}} \int |\mathcal{M}|^2 (2\pi)^4 \delta^4(p_1 + p_2 - p_3 - p_4 \ldots p_n) \times \prod_{j=3}^{n} \frac{1}{2\sqrt{\vec{p}_j^{\,2} + m_j^2}} \frac{d^3\vec{p}_j}{(2\pi)^3}. \tag{3.14}$$

In the case of 2-to-2 scattering, $1 + 2 \to 3 + 4$, evaluated in the center-of-momentum frame, the differential cross section per solid angle reduces to

$$\frac{d\sigma}{d\Omega} = \frac{S\,|\mathcal{M}|^2}{64\pi^2 (E_1 + E_2)^2} \frac{|\vec{p}_f|}{|\vec{p}_i|}, \tag{3.15}$$

where $|\vec{p}_i|$ is the magnitude of the momentum of either incoming particle, and $|\vec{p}_f|$ is the magnitude of the momentum of either outgoing particle.

Let's use this expression to estimate the cross section for Compton scattering, $\gamma e^- \to \gamma e^-$. In Sec. 3.5, we estimated the squared amplitude for this process to be

$|\mathcal{M}|^2 \sim g_2^4 \sin^4\theta_W$. Combining this with Eq. 3.15, we arrive at $\sigma \sim g_2^4 \sin^4\theta_W / 16\pi E_{\rm tot}^2$, where $E_{\rm tot} \equiv E_\gamma + E_e$. This result is not very different from the more carefully calculated answer of $\sigma \approx (g_2^4 \sin^4\theta_W / 8\pi E_{\rm tot}^2) \ln(E_{\rm tot}^2/m_e^2)$.

Additional reading

For a highly approachable textbook on particle physics, I recommend Griffiths' *Introduction to Elementary Particles*. For a somewhat more advanced (but thematically similar) treatment, Halzen and Martin's *Quarks and Leptons* has long been the standard, and for good reason. I'm also fond of Thomson's *Modern Particle Physics*, Larkoski's *Elementary Particle Physics*, and Palash's *An Introductory Course of Particle Physics*. When it comes to textbooks on the subject of quantum field theory, many people have strong opinions and preferences. An important classic is Peskin and Schroeder's *An Introduction to Quantum Field Theory*, which still holds up. I also like Schwartz's *Quantum Field Theory and the Standard Model* quite a lot. Others prefer Srednicki's *Quantum Field Theory* or Weinberg's *Quantum Theory of Fields*, both of which are excellent texts. If you're looking for an introduction to some of the concepts behind quantum field theory, with less emphasis on the technical aspects or calculations, I recommend Lancaster's *Quantum Field Theory for the Gifted Amateur*.

Problems

3.1 Write down all of the tree-level QED Feynman diagrams for the processes $e^+e^- \to \mu^+\mu^-$, $e^-\mu^- \to e^-\mu^-$, $\mu^-\mu^- \to \mu^-\mu^-$, and $\mu^+\mu^- \to \mu^+\mu^-$. Hint: There are one, one, two, and two tree-level diagrams for these processes, respectively.

3.2 Write down all 18 of the one-loop diagrams involving only photons, electrons, and positrons for the process $e^-e^- \to e^-e^-$.

3.3 Considering only the quarks u, d, s, and their antiparticles, write down all of the mesons, baryons, and antibaryons they can form. What is the total electric charge of each of these hadrons?

3.4 Consider quark-quark elastic scattering, $qq \to qq$. What are the main Feynman diagrams that describe this process? Roughly estimate the ratio of the squared amplitudes that result from gluon exchange and from photon exchange. Evaluate this ratio at an energy of 1 GeV and 100 GeV.

3.5 Repeat the previous question, making estimates for the squared amplitudes for electron-electron and neutrino-neutrino elastic scattering.

3.6 Show that the spacetime interval $t^2 - x^2 - y^2 - z^2$ is a Lorentz invariant.

3.7 Show that Eq. 3.11 follows from Eq. 3.10.

3.8 The squared amplitude for neutral pion decay, $\pi^0 \to \gamma\gamma$, is given by:

$$|\mathcal{M}|^2 = \frac{N_c^2 \alpha^2 m_\pi^4}{72\pi^2 F_\pi^2}, \tag{3.16}$$

where $N_c = 3$ is the number of QCD colors, and $F_\pi \approx 0.13\,\text{GeV}$ is a quantity known as the pion decay constant. Calculate the rate for this process. Using the fact that this is the main decay channel for a neutral pion, what is the lifetime of this particle?

3.9 Using the entries of the CKM matrix, calculate the ratio of the $W^\pm$ boson's decay rates to $u\bar{d}$, $u\bar{s}$, and $u\bar{b}$.

3.10 Treating the electron as massless (a good approximation in this case), the squared amplitude for muon decay, $\mu^- \to e^- \nu_\mu \bar{\nu}_e$, is given by

$$|\mathcal{M}|^2 = \frac{g_2^4 m_\mu^2}{m_W^4} \, |\vec{p}_{\bar{\nu}_e}| \, (m_\mu - 2|\vec{p}_{\bar{\nu}_e}|), \tag{3.17}$$

where $\vec{p}_{\bar{\nu}_e}$ is the momentum of the outgoing antineutrino. Use Eq. 3.11 to calculate the rate for this process. Hint: Start by integrating over $\vec{p}_{\nu_\mu}$. Next write $d^3 p_{\bar{\nu}_e} = |\vec{p}_{\bar{\nu}_e}|^2 d|p_{\bar{\nu}_e}| \, \sin\theta \, d\theta \, d\phi$, and make the substitution $u \equiv |\vec{p}_{\nu_\mu} + \vec{p}_{\bar{\nu}_e}|$.

3.11 Treating the electron as massless (again, a good approximation), the squared amplitude for the process $e^- \nu_\mu \to \nu_e \mu^-$ is given by

$$|\mathcal{M}|^2 = 8 g_2^4 \left(\frac{E}{m_W} \right)^4 \left[1 - \left(\frac{m_\mu}{2E} \right)^2 \right], \tag{3.18}$$

where E is the energy of the incoming positron or neutrino. Use Eq. 3.15 to find the total cross section for this process.

3.12 What is the relationship between the lifetime of an unstable particle and its half-life?

3.13 Show that Eq. 3.14 follows from Eq. 3.13. Hint: Make use of the following property of the delta function: $\delta(g(x)) = [g'(x_0)]^{-1} \, \delta(x - x_0)$.

4

Thermodynamics for Cosmologists

The subject of thermodynamics is integral to many of the questions asked by cosmologists, especially in regard to understanding the early universe. In this chapter, I review concepts such as thermodynamic equilibrium and entropy, as commonly encountered in cosmological applications. I then move on to discuss departures from equilibrium, including the derivation and uses of various forms of the Boltzmann equation.

4.1 Equilibrium thermodynamics

As our universe expanded, it also cooled. In many respects, cosmic history is the story of how our universe gradually transitioned from the incredibly hot and dense state known as the Big Bang, to the cool and dilute space that we find ourselves living in today. In the very early universe, particles typically interacted frequently enough with one another to maintain a common temperature, and to repeatedly transition back and forth between various species. Under such conditions, a universe can be described as existing in a state of thermodynamic equilibrium.

There are two distinct kinds of equilibria that we will be concerned about in this chapter:

- *Kinetic equilibrium* is reached when reactions that exchange energy and momenta between particles occur frequently compared to the timescales under which conditions are otherwise changing.
- *Chemical equilibrium* is reached when reactions that transform particles into particles of other species occur frequently compared to the timescales under which conditions are otherwise changing.

In most cosmological applications, equilibrium is reached when the reaction rate in question is greater than the rate of Hubble expansion.

Consider a collection of particles of mass m that have been scattering or otherwise interacting with one another long enough to reach both kinetic and chemical equilibrium. In this equilibrium state, the distribution of the particles' momenta is determined by their temperature, T:

$$\frac{dn}{d^3\vec{p}} = \frac{g f(E)}{(2\pi)^3} = \frac{g}{(2\pi)^3} \frac{1}{e^{(E-\mu)/T} \pm 1}, \tag{4.1}$$

Table 4.1 The number of internal degrees of freedom for each of the particles in the Standard Model. The entire Standard Model contains 90 internal degrees of freedom in fermions and 28 in bosons.

	Flavors	Particle+Antiparticle	Colors	g_s	g (total)
Quarks (d, u, s, c, b, t)	6	2	3	2	72
Charged Leptons (e, μ, τ)	3	2	1	2	12
Neutrinos (ν_e, ν_μ, ν_τ)	3	2	1	1	6
Gluons (g)	1	1	8	2	16
Photon (γ)	1	1	1	2	2
$W^\pm$	1	2	1	3	6
Z	1	1	1	3	3
Higgs Boson (h)	1	1	1	1	1

where $f(E) = (e^{(E-\mu)/T} \pm 1)^{-1}$ is the phase space distribution function (also known as the phase space occupancy), $E^2 = \vec{p}^{\,2} + m^2$, and the plus sign in the denominator applies to the case of fermions, while the minus is appropriate for bosons. The symbol μ is the chemical potential of the particle species in question, which, as we will see later, will be useful for keeping track of and imposing various conservation laws. Lastly, g counts the number of internal degrees of freedom of the particle species (see Table 4.1). This includes all of the possible spin states, as well as states with different charges or other quantum numbers. The number of spin states associated with a particle species of mass m and spin s is given by

$$g_s = \begin{cases} 2s+1, & m > 0, \\ 2, & m = 0,\ s > 0, \\ 1, & s = 0. \end{cases} \tag{4.2}$$

For a concrete example, the photon has two spin states, $+1$ and -1. Since it has no charges or other quantum numbers to count, and is its own antiparticle, its total number of internal degrees of freedom is simply $g_\gamma = 2$. In contrast, each quark species (including antiquarks) has $g_q = 2 \times 6 = 12$ degrees of freedom, where two is the number of spin states ($+1/2$ and $-1/2$) and six is the number of different colors and anticolors the quark or antiquark could carry. Considering all six flavors, the quarks of the Standard Model have a total of 72 degrees of freedom.

The distribution given in Eq. 4.1 describes the number of particles per volume, in a given three-dimensional volume of momentum space. If we were instead interested in the distribution of these particles' energies, we can relate

$$\begin{aligned} d^3\vec{p} &= 4\pi\, \vec{p}^{\,2} d|\vec{p}| \\ &= 4\pi\, (E^2 - m^2)^{1/2}\, E\, dE, \end{aligned} \tag{4.3}$$

leading to

$$\frac{dn}{dE} = \frac{g}{2\pi^2} \frac{E(E^2 - m^2)^{1/2}}{e^{(E-\mu)/T} \pm 1}. \tag{4.4}$$

By integrating either Eq. 4.1 or 4.4, we can determine the number density of a particle species in chemical equilibrium. On similar grounds, we can obtain the corresponding

energy density by multiplying the distribution by E before integrating, or the pressure by first multiplying by $\vec{p}^{\,2}/3E = (E^2 - m^2)/3E$:

$$n = \frac{g}{2\pi^2}\int_m^\infty \frac{(E^2-m^2)^{1/2}}{e^{(E-\mu)/T}\pm 1}EdE, \tag{4.5}$$

$$\rho = \frac{g}{2\pi^2}\int_m^\infty \frac{(E^2-m^2)^{1/2}}{e^{(E-\mu)/T}\pm 1}E^2dE,$$

$$p = \frac{g}{6\pi^2}\int_m^\infty \frac{(E^2-m^2)^{3/2}}{e^{(E-\mu)/T}\pm 1}dE.$$

Some limiting cases of these expressions are useful to consider. First, if the particles in question are highly relativistic and have a negligible chemical potential ($T \gg m, \mu$), these integrals reduce to

$$n = \begin{cases} \dfrac{\zeta(3)gT^3}{\pi^2} & \text{(Bose-Einstein)}, \\ \dfrac{3}{4}\dfrac{\zeta(3)gT^3}{\pi^2} & \text{(Fermi-Dirac)}, \end{cases} \tag{4.6}$$

$$\rho = \begin{cases} \dfrac{\pi^2 gT^4}{30} & \text{(Bose-Einstein)}, \\ \dfrac{7}{8}\dfrac{\pi^2 gT^4}{30} & \text{(Fermi-Dirac)}, \end{cases}$$

$$p = \begin{cases} \dfrac{\pi^2 gT^4}{90} & \text{(Bose-Einstein)}, \\ \dfrac{7}{8}\dfrac{\pi^2 gT^4}{90} & \text{(Fermi-Dirac)}, \end{cases}$$

where $\zeta(3) \approx 1.20206$ is the Riemann zeta function evaluated at 3. We can take the ratio of ρ and n to determine the mean energy per particle as a function of temperature in this limit:

$$\langle E\rangle \equiv \frac{\rho}{n} = \begin{cases} \dfrac{\pi^4 T}{30\zeta(3)} \approx 2.70\,T & \text{(Bose-Einstein)}, \\ \dfrac{7}{6}\dfrac{\pi^4 T}{30\zeta(3)} \approx 3.15\,T & \text{(Fermi-Dirac)}. \end{cases} \tag{4.7}$$

In the opposite, nonrelativistic limit ($T \ll m$), the integrals instead reduce to the well-known Maxwell-Boltzmann distribution:

$$n = g\left(\frac{mT}{2\pi}\right)^{3/2} e^{-(m-\mu)/T}, \tag{4.8}$$

$$\rho = g\,m^{5/2}\left(\frac{T}{2\pi}\right)^{3/2} e^{-(m-\mu)/T},$$

$$p = g\, T^{5/2} \left(\frac{m}{2\pi}\right)^{3/2} e^{-(m-\mu)/T}.$$

Notice that in the case of nonrelativistic particles, $\rho = mn$ and $p = nT$. Since $nT \ll \rho$, the pressure associated with nonrelativistic matter can usually be safely neglected when evaluating the Friedmann equations.

In practice, a cosmologist rarely considers just one particle species at a time. The energy density of an ensemble of species can be written as a function of temperature in terms of the quantity, $g_\star$:

$$\rho = \sum_i \rho_i \equiv \frac{\pi^2 g_\star(T) T^4}{30}, \tag{4.9}$$

where the sum is over all of the particle species of interest, and T is defined as the temperature of the photons in the bath. This expression can be thought of as the definition of $g_\star$, which is generally referred to as the number of relativistic degrees of freedom.

If all of the particle species in question are highly relativistic, we can use Eq. 4.6 to write their total energy density as

$$\begin{aligned} \rho &= \sum_{i=\text{bosons}} \frac{\pi^2 g_i T_i^4}{30} + \sum_{i=\text{fermions}} \frac{7}{8} \frac{\pi^2 g_i T_i^4}{30} \\ &= \frac{\pi^2 T^4}{30} \left[\sum_{i=\text{bosons}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{i=\text{fermions}} g_i \left(\frac{T_i}{T}\right)^4 \right], \end{aligned} \tag{4.10}$$

where T is the photon temperature, and T_i is the temperature of species i. Comparing this to the definition of $g_\star$ in Eq. 4.9, we see that if all of the particle species are in kinetic equilibrium ($T_i = T$, for all i) and highly relativistic, then $g_\star$ is simply equal to the total number of bosonic degrees of freedom, plus 7/8 times the total number of fermionic degrees of freedom. For any species that is not highly relativistic, however, we cannot use Eq. 4.6, but instead must carry out the integral given in Eq. 4.5, resulting in a suppressed contribution to $g_\star$.

In many instances, it can be relatively straightforward for calculate $g_\star$. Let's consider a few examples.

- In the very early universe ($T \gg 100$ GeV), all of the particle species described by the Standard Model were present and maintained at a common temperature. When we add up the total number of bosonic degrees of freedom, plus 7/8 times the total number of fermionic degrees of freedom contained in the Standard Model, we get $g_\star = 28 + (7/8) \times 90 = 106.75$.
- At $T \sim 1 - 100$ MeV, photons, electrons, neutrinos, and their antiparticles were all present and maintained at a common temperature. During this period, $g_\star \approx 2 + (7/8) \times 4 + (7/8) \times 6 = 10.75$.
- At temperatures below $T \sim m_e$, photons and neutrinos were both present, but were no longer maintained at the same temperature. As we will see later in this chapter, the photon and neutrino temperatures during this era were related as $T_\nu = (4/11)^{1/3}\, T$, resulting in $g_\star \approx 2 + (7/8) \times 6 \times (4/11)^{4/3} \approx 3.36$.

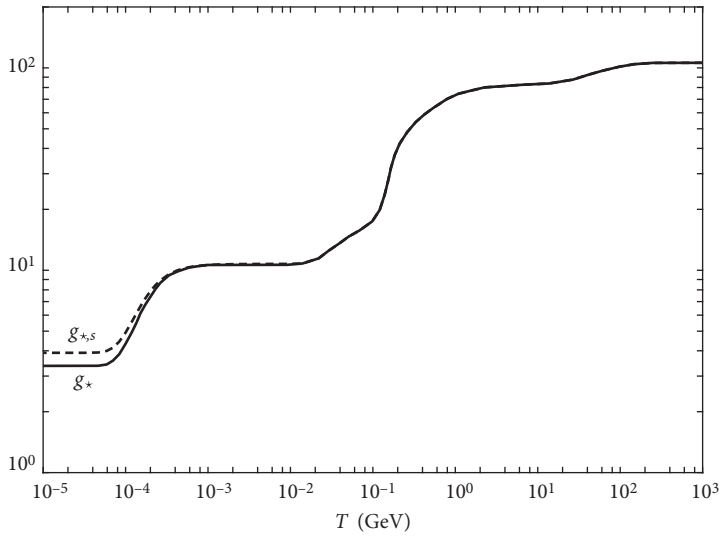


Figure 4.1 The number of relativistic degrees of freedom, $g_\star$, and the number of relativistic degrees of freedom in entropy, $g_{\star,S}$, in the Standard Model as a function of the photon temperature.

In Figure 4.1, the value of $g_\star$ is shown as a function of temperature. At most temperatures, the calculation of $g_\star$ is not particularly complicated. The exceptions are those temperatures between a few hundred MeV and a few GeV, near the QCD phase transition. The rapid increase in $g_\star$ that takes place in this temperature range is the result of the many mesonic and baryonic states that contribute to the energy density.

4.2 Entropy

Another useful quantity in considering the thermal history of our universe is the entropy density:

$$s \equiv \frac{S}{V} = \frac{\rho + p}{T}, \tag{4.11}$$

where the last step follows from the relationship between the change in the internal energy of a system and the change in its entropy and volume, $dE = TdS - pdV$.

The entropy density can be written in terms of the photon temperature and a quantity known as the number of relativistic degrees of freedom in entropy, $g_{\star,S}$:

$$s = \sum_i s_i \equiv \frac{2\pi^2 g_{\star,S}(T)\, T^3}{45}, \tag{4.12}$$

where the sum is performed over all of the particle species in the bath. This relationship can be thought of as the definition of $g_{\star,S}$. In most cases of interest, the entropy density is dominated by the relativistic particles that are present, for which $p = \rho/3$ and thus $s = (4/3)\rho/T$. In this limit, we can write

$$s = \frac{4}{3}\frac{\rho}{T} = \frac{4}{3}\left[\sum_{i=\text{bosons}} \frac{\pi^2 g_i T_i^4}{30}\frac{1}{T_i} + \sum_{i=\text{fermions}} \frac{7}{8}\frac{\pi^2 g_i T_i^4}{30}\frac{1}{T_i}\right] \tag{4.13}$$

$$= \frac{2\pi^2 T^3}{45}\left[\sum_{i=\text{bosons}} g_i\left(\frac{T_i}{T}\right)^3 + \frac{7}{8}\sum_{i=\text{fermions}} g_i\left(\frac{T_i}{T}\right)^3\right].$$

Comparing this to Eq. 4.12, we see that for the case of relativistic particles, $g_{\star,S}$ is equal to the quantity in the square brackets. Recall that T is defined to be the temperature of the photons in the thermal bath, while T_i is the temperature of species i.

Note that the quantities $g_\star$ and $g_{\star,S}$ are similar, except that the terms contributing to $g_{\star,S}$ scale as $(T_i/T)^3$ instead of $(T_i/T)^4$. If all of the particle species in question are maintained at a common temperature, these two quantities will be equal to one another, $g_\star \simeq g_{\star,S}$. For the particles that make up the Standard Model, this was the case in the early universe at temperatures above $T \sim$ MeV. At lower temperatures, however, the neutrinos were cooler than the rest of the bath, leading to $g_{\star,S} > g_\star$. The value of $g_{\star,S}$ as calculated in the Standard Model is shown as a function of the photon temperature in Figure 4.1.

Entropy is a useful concept in cosmology primarily because it is conserved in thermal equilibrium. We are all used to the second law of thermodynamics requiring entropy to increase with time, driving things toward a state of equilibrium. For a system that is already in equilibrium, however, the total entropy cannot increase any further. In such situations, the second law simply requires the total entropy to remain constant. More precisely, so long as equilibrium is maintained, the total amount of entropy within a given comoving volume will be conserved as space expands.

Applying the conservation of entropy to the expanding universe, the entropy density scales with the inverse volume, or three powers of the scale factor:

$$S = s\,V \propto s\,a^3 \rightarrow s \propto a^{-3}. \tag{4.14}$$

Combining this with Eq. 4.12, we conclude that the quantity $g_{\star,S}\,T^3\,a^3$ remains constant as space expands. Equivalently, the temperature evolves as follows, in terms of the scale factor and $g_{\star,S}$:

$$T \propto \frac{1}{a\,g_{\star,S}^{1/3}}. \tag{4.15}$$

As space expands, the temperature generally drops as $T \propto a^{-1}$, except when the temperature is near the mass of one or more of the particle species that are present in the thermal bath. When this occurs, those massive particles begin to be destroyed more often than they are created, reducing their abundance and transferring their energy into lighter particle species. This phenomenon is accounted for in the change of $g_{\star,S}$. This kind of transfer of energy from a massive species to lighter ones is sometimes referred to as an "entropy dump." We will explore one such example of an entropy dump in Sec. 4.5.

4.3 Chemical potentials

The chemical potential of a given particle species is defined as the amount of energy that is absorbed or released whenever a member of that species is created or destroyed (at constant entropy and volume). For a collection of particle species that are in chemical equilibrium, their chemical potentials can be related to one another. Consider, for example, an electron-positron pair, which can annihilate to become a muon-antimuon pair. If such interactions occur frequently enough such that chemical equilibrium is maintained, we can

impose the condition $\mu_{e^-} + \mu_{e^+} = \mu_{\mu^-} + \mu_{\mu^+}$. Expressions such as these effectively encode information pertaining to the conservation of electric charge, lepton number, and other quantum numbers.

As another example, consider a collection of electrons, positrons, and photons, all in chemical equilibrium. Since photons can interact among themselves in a way that changes their number (such as $2\gamma \rightarrow 3\gamma$, for example), their chemical potential must be zero: $2\mu_\gamma = 3\mu_\gamma \longrightarrow \mu_\gamma = 0$. This is simply a reflection of the fact that photons carry no conserved quantum numbers. Furthermore, since electron-positron pairs can annihilate to produce a pair of photons, we can write $\mu_{e^-} + \mu_{e^+} = 2\mu_\gamma = 0$, and thus $\mu_{e-} = -\mu_{e^+}$.

Using Eq. 4.5, we can write the difference between the number densities of electrons and positrons as follows:

$$n_{e^-} - n_{e^+} = \frac{g_{e^-}}{2\pi^2}\int_{m_e}^{\infty} \frac{(E^2 - m_e^2)^{1/2}}{e^{(E-\mu_{e^-})/T} + 1} EdE - \frac{g_{e^+}}{2\pi^2}\int_{m_e}^{\infty} \frac{(E^2 - m_e^2)^{1/2}}{e^{(E-\mu_{e^+})/T} + 1} EdE \tag{4.16}$$

$$= \frac{1}{\pi^2}\left[\int_{m_e}^{\infty} \frac{(E^2 - m_e^2)^{1/2}\, E\, dE}{e^{(E-\mu_{e^-})/T} + 1} - \int_{m_e}^{\infty} \frac{(E^2 - m_e^2)^{1/2}\, E\, dE}{e^{(E+\mu_{e^-})/T} + 1}\right],$$

where in the last step I have used the fact that $g_{e^-} = g_{e^+} = 2$ and $\mu_{e+} = -\mu_{e^-}$. If the chemical potential is small compared to the temperature, $|\mu_{e^-}| \ll T$, the two terms will be approximately equal in magnitude, yielding $|n_{e^-} - n_{e^+}| \ll n_{e^-}, n_{e^+}$. In other words, there will be approximately equal numbers of electrons and positrons present in the thermal bath. In the opposite limit of a large and positive chemical potential for the electrons, $\mu_{e^-} \gg T$, we instead find that there will be far more electrons than positrons, $n_{e^-} \gg n_{e^+}$. Physically speaking, this is because most of the positrons have been annihilated, leaving behind a much greater abundance of electrons.

In the high- and low-temperature limits, Eq. 4.16 reduces to

$$n_{e^-} - n_{e^+} \approx \begin{cases} \dfrac{\mu_{e^-} T^2}{3} + \dfrac{\mu_{e^-}^3}{3\pi^2}, & T \gg m_e, \\ 4\left(\dfrac{m_e T}{2\pi}\right)^{3/2} \sinh\left(\dfrac{\mu_{e^-}}{T}\right) e^{-m_e/T}, & T \ll m_e. \end{cases} \tag{4.17}$$

To develop some intuition for the role that chemical potentials can play, consider an example of how the number densities of electrons and positrons would have evolved in the early universe if $\mu_{e^-} = 100$ keV (and thus $\mu_{e^+} = -100$ keV). At very early times, when $T \gg m_e, |\mu_{e^-}|$, the number densities of electrons and positrons would each be very similar, $n_{e^-} \approx n_{e^+} \approx 3\zeta(3)T^3/2\pi^2$. At $T \sim 10$ MeV, for example, this yields $n_{e^-} \approx n_{e^+} \approx 3\zeta(3)T^3/2\pi^2 \approx 2 \times 10^{-7}$ GeV^3.[1] Comparing this to $n_{e^-} - n_{e^+} \approx 3 \times 10^{-9}$ GeV$^3 \ll n_{e^-}, n_{e^+}$, it is clear that the chemical potential has only a small impact on the distributions of these particles. When the temperature drops to $T \sim \mu_{e^-}$, however, the chemical potential begins to play an important role. In particular, at around $T \sim \mu_{e^-}$, $n_{e^-} - n_{e^+}$ grows to become the same order of magnitude as n_{e^-}, and much larger than n_{e^+}. In other words, the number

[1] If you're not in the habit of using natural units, it might seem strange to have a number density in units of GeV^3. To convert this to units of length^{-3}, one simply multiplies by the quantity $(\hbar c)^{-3}$, which is equal to 1 in natural units.

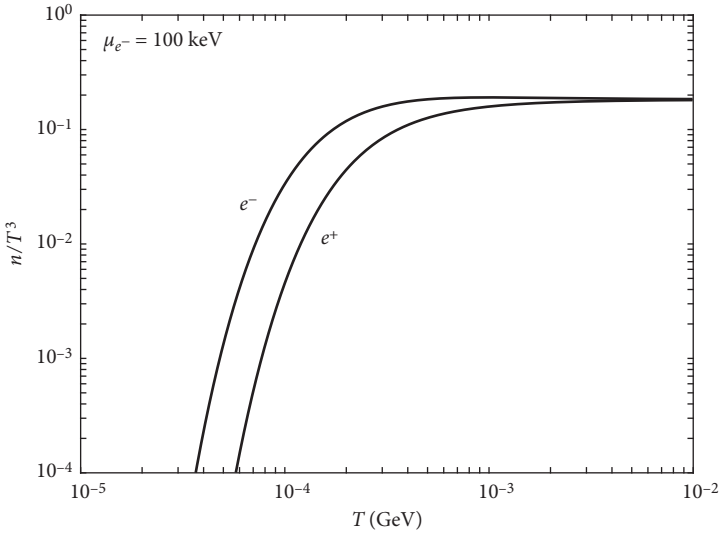


Figure 4.2 The number densities of electrons and positrons as a function of temperature, adopting a chemical potential of $\mu_{e^-} = 100$ keV (and therefore $\mu_{e^+} = -100$ keV).

density of electrons begins to significantly exceed that of positrons at this time. By $T \sim 10$ keV, $n_{e^-}/n_{e^+} \approx 5 \times 10^8$, reflecting the fact that the vast majority of the positrons have been annihilated away. This leaves behind a much larger abundance of electrons which, in the absence of positrons, cannot be further destroyed.

4.4 Departures from equilibrium

A state of equilibrium is reached whenever the relevant interactions between particles occur at a much faster rate than the rate at which the environment is changing. In cosmology, this typically means that the rate of one or more scattering processes exceeds the rate of Hubble expansion:

$$\Gamma = n\,\sigma\,v \gtrsim H \qquad \text{(equilibrium condition)}, \tag{4.18}$$

where n is the number density of target particles, σ is the scattering cross section, and v is the relative velocity of the particles. As we will see, it is often the case that a given collection of particles will be in equilibrium during one period of cosmic history, only to decouple and fall out of equilibrium at a later time, when the above condition is no longer satisfied.

As a first example, consider the scattering between Standard Model fermions. This could include both the elastic scattering of such particles, as well as inelastic processes, in which particles can be destroyed and replaced by particles of another species. While elastic scattering can drive a collection of particles to approach kinetic equilibrium, species-changing processes are important if the particles are to reach and maintain chemical equilibrium.

At temperatures well above the electroweak phase transition, $T \gg T_{\rm EWPT} \sim 200$ GeV, fermion-antifermion pairs can annihilate and transform into other fermionic species through the exchange of massless gauge bosons. The cross section for this process, up to order 1 factors, is given by $\sigma \sim \pi\alpha^2/E^2 = g_2^4 \sin^4\theta_W/16\pi E^2$, where E is the energy

of one of the fermions in the center-of-momentum frame. Using the number density of relativistic fermions given in Eq. 4.6, and substituting $E \sim 3.15\, T$ (see Eq. 4.7), we find that the scattering rate at high temperatures is approximately given by $\Gamma \sim 10^{-4}\, T$. Comparing this to the Hubble rate as determined by the first Friedmann equation, $H = (8\pi\rho/3M_{\rm Pl}^2)^{1/2}$, and using Eq. 4.9 to determine the total energy density, we find that both chemical and kinetic equilibrium should be maintained so long as

$$\frac{\Gamma}{H} \sim \frac{10^{-4}\, T}{(8\pi^3 g_\star(T) T^4/90 M_{\rm Pl}^2)^{1/2}} \sim \frac{10^{14}\,{\rm GeV}}{T} \gtrsim 1. \tag{4.19}$$

In other words, at temperatures between $T \sim 10^{14}$ GeV and $T \sim T_{\rm EWPT}$, the scattering rate exceeded the rate of Hubble expansion, maintaining these particles (and all of the Standard Model particles species) in a state of both kinetic and chemical equilibrium.

At temperatures below the electroweak phase transition, we have to treat electrically charged fermions differently from neutrinos. Interactions with massless photons involve cross sections of the order of $\sigma \sim \pi\alpha^2/E^2 \sim g_2^4 \sin^4\theta_W/16\pi E^2$, while those that rely on the massive weak gauge bosons provide only $\sigma \sim G_F^2 E^2 \sim g_2^4 E^2/32 m_W^4$. Comparing the rates for various scattering processes to the Hubble rate, we find several instances in which a particle species fell out of chemical or kinetic equilibrium as the universe expanded and cooled. Some examples include the following:

- Photons and electrons decoupled and departed from kinetic equilibrium with each other at a temperature of $T \approx 0.3$ eV, about 380,000 years after the Big Bang. This occurred not only because of the reduction in the number density of electrons, but also because of the reduction in the number density of *free* electrons, as such particles in this era were becoming bound to protons, forming electrically neutral atoms. The photons that decoupled during this transition now make up the cosmic microwave background. This transition will be described in some detail in Chapter 6.
- Protons and neutrons can transform into each another through processes involving the weak force, $p\, e^- \leftrightarrow n\, \nu_e$, $p\, \bar{\nu}_e \leftrightarrow n\, e^+$. The rate for such processes exceeded that of Hubble expansion only when $T \gtrsim$ MeV, after which protons and neutrons were no longer maintained in chemical equilibrium. Interactions involving the strong force kept these particles in kinetic equilibrium up to much later times, however. We will discuss the process of proton-neutron freeze out in more detail in Chapter 8.
- Neutrinos were kept in both kinetic and chemical equilibrium with the rest of the Standard Model bath through interactions involving the weak force until $T \sim$ MeV, at which point they decoupled. This transition will be the topic of the following section.

4.5 Neutrino decoupling

Neutrinos were maintained in equilibrium with the rest of the Standard Model bath in the early universe through the exchange of $W^\pm$ and Z bosons. Focusing on the case of $T \ll m_\mu$, the processes involved are $\nu\bar{\nu} \leftrightarrow e^+e^-$, via either s-channel Z exchange or t-channel $W^\pm$ exchange, as well as $\nu e^+ \leftrightarrow \nu e^+$ and $\bar{\nu} e^- \leftrightarrow \bar{\nu} e^-$ through s-channel $W^\pm$ exchange. The cross section for any of these processes, averaged over the energy distribution of particles at a

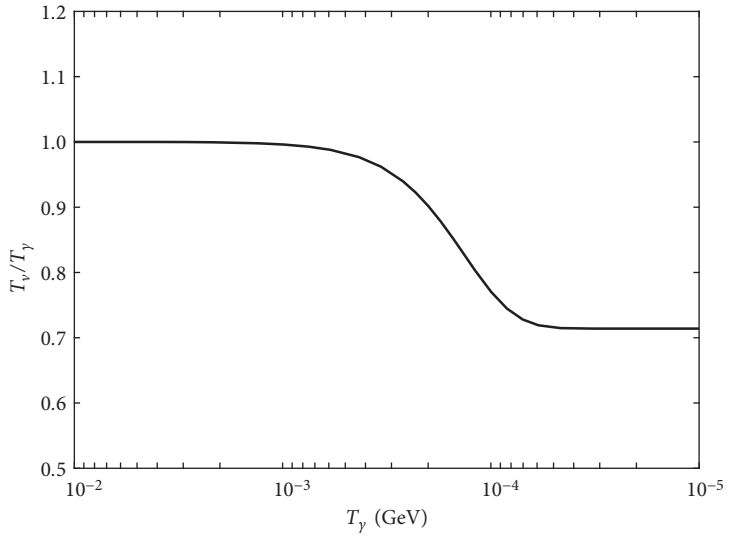


Figure 4.3 The ratio of the neutrino and photon temperatures during the era of electron-positron annihilation.

temperature T, is roughly given by $\sigma \approx G_F^2 T^2 = g_2^4 T^2/32 m_W^4$. Note that while high rates of pair production and annihilation could maintain the neutrinos in chemical equilibrium, any of the processes listed above could potentially keep these particles in kinetic equilibrium with the electrons and photons that made up the rest of the Standard Model bath during this era.

Using the first Friedmann equation and the number density of neutrinos or electrons from Eq. 4.6, the ratio of the relevant scattering rates to that of Hubble expansion is given by

$$\begin{aligned} \frac{\Gamma}{H} &= \frac{n_{e,\nu}\,\sigma}{(8\pi\rho/3M_{\rm Pl}^2)^{1/2}} \\ &\sim \left(\frac{10.75}{g_\star(T)}\right)^{1/2} \left(\frac{T}{3\,\text{MeV}}\right)^3 . \end{aligned} \tag{4.20}$$

From this exercise, we learn that at temperatures above a few MeV, neutrinos interacted rapidly enough with electrons to remain in both kinetic and chemical equilibrium. Around $T \sim 3$ MeV, however, the Hubble rate overtook the rate of neutrino scattering, causing the neutrinos to decouple from the rest of the thermal bath.

Although the temperature of the neutrinos at the time of their decoupling was equal to that of the electrons and photons, this did not remain true for long. As the temperature of the photons and electrons dropped below the electron mass, most of the electrons and positrons annihilated, with their number density falling as $n_e \propto T^{3/2}\, e^{-m_e/T}$ (see Eq. 4.8). These annihilations heated the photons, causing T (and T_e) to increase and thus to exceed the value of T_ν. To determine how the temperatures of these two decoupled sets of particles evolved, we can make use of entropy, which is conserved for the photons and electrons, and is independently conserved for the neutrinos. Note that while the photons and electrons were not in equilibrium with the neutrinos at this time, the members of each of these particle species were still described by an equilibrium distribution (with $T = T_e > T_\nu$), thus allowing us to safely apply the conservation of entropy.

From Eq. 4.15, we find that, in equilibrium, the temperature evolves as $T \propto a^{-1}\, g_{\star,S}^{-1/3}$. For the neutrinos, $g_{\star,S}^{\nu}$ was fixed at $(7/8) \times 6$ throughout this period of time. For the electrons and photons, in contrast, $g_{\star,S}^{\gamma}$ decreased from $(7/8) \times 4 + 2$ to simply 2 as the electrons and positrons annihilated and disappeared from the bath. From this, we can determine the temperatures of the photons and neutrinos after the era of electron-positron annihilation:

$$\begin{aligned} T &= T_{\rm dc} \times \left(\frac{a_{\rm dc}}{a}\right) \left(\frac{g_{\star,S}^{\gamma}(T_{\rm dc})}{g_{\star,S}^{\gamma}(T)}\right)^{1/3} \\ &= T_{\rm dc} \times \left(\frac{a_{\rm dc}}{a}\right) \left(\frac{\frac{7}{8} \times 4 + 2}{2}\right)^{1/3} \end{aligned} \tag{4.21}$$

and

$$\begin{aligned} T_{\nu} &= T_{\rm dc} \times \left(\frac{a_{\rm dc}}{a}\right) \left(\frac{g_{\star,S}^{\nu}(T_{\rm dc})}{g_{\star,S}^{\nu}(T_{\nu})}\right)^{1/3} \\ &= T_{\rm dc} \times \left(\frac{a_{\rm dc}}{a}\right) \left(\frac{\frac{7}{8} \times 6}{\frac{7}{8} \times 6}\right)^{1/3}, \end{aligned} \tag{4.22}$$

where the subscript "dc" denotes quantities evaluated at the time of neutrino decoupling. Combining these results, we find the following relationship between the temperatures of the photons and the neutrinos after the completion of electron-positron annihilation, $T \ll m_e$:

$$T_{\nu} = T \left(\frac{g_{\star,S}^{\gamma}(T)}{g_{\star,S}^{\gamma}(T_{\rm dc})}\right)^{1/3} \xrightarrow{T \ll m_e} T \times \left(\frac{4}{11}\right)^{1/3} \approx 0.71\, T. \tag{4.23}$$

This kind of heat transfer is a classic example of an entropy dump.

After the era of electron-positron annihilation, the value of T/T_{ν} remained fixed at the value presented in Eq. 4.23, until the neutrinos became nonrelativistic, $T_{\nu} \sim m_{\nu}$, at around a redshift of $z \sim 600 \times (m_{\nu}/0.1\,{\rm eV})$. If the neutrinos had been massless, the temperatures of the cosmic neutrino and cosmic microwave backgrounds would have been simply related according to this expression. In particular, whereas the temperature of the cosmic microwave background has been measured to be $T_{\rm CMB} = 2.725\,{\rm K}$, the temperature of a massless cosmic neutrino background would have been somewhat cooler, $T_{\nu} = T_{\rm CMB} \times (4/11)^{1/3} = 1.945\,{\rm K}$.

4.6 Evolution after decoupling

A collection of particles will come to be described by the distributions described in Sec. 4.1 when they interact enough to reach both kinetic and chemical equilibrium. In this section, we'll explore what happens to a population of stable particles after they have decoupled and departed from their equilibrium state.

As a first example, let's consider the case of a massless or otherwise highly relativistic particle species. Up until the time of decoupling, these particles will be described by an equilibrium distribution, as given in Eq. 4.4. After decoupling, the energy of an individual relativistic particle evolves due to cosmological redshift, decreasing in proportion to the

inverse of the scale factor, $E \propto a^{-1}$. In addition, since the total number of decoupled particles is conserved, their number density drops as the volume increases, $n \propto a^{-3}$.

To see what effect these changes have on the distribution of these particles, consider the number density of the particles within a narrow energy interval, E to $E + \Delta E$, evaluated at a time after decoupling. After accounting for cosmological redshift, these particles have energies in the range of $E \times (a/a_{\rm dc})$ to $(E + \Delta E) \times (a/a_{\rm dc})$, where $a_{\rm dc}$ is the scale factor at decoupling. Combining this with the fact that the universe has grown in volume by a factor of $(a/a_{\rm dc})^3$ since the time of decoupling, we can use Eq. 4.4 to write the number density of particles in this energy range:

$$\begin{aligned} \Delta n &= \frac{g}{2\pi^2} \frac{(E\,a/a_{\rm dc})^2\,(\Delta E\,a/a_{\rm dc})}{e^{(E\,a/a_{\rm dc})/T_{\rm dc}} \pm 1} \times \left(\frac{a_{\rm dc}}{a}\right)^3 \\ &= \frac{g}{2\pi^2} \frac{E^2\,\Delta E}{e^{E/T} \pm 1}, \end{aligned} \tag{4.24}$$

where in the last line I have substituted $(a_{\rm dc}/a)\,T_{\rm dc} \to T$. Comparing this to Eq. 4.4, this exercise shows us that a collection of relativistic particles that starts out in an equilibrium distribution will remain in an equilibrium distribution if those particles are left undisturbed as the universe expands. The distribution only changes in the sense that its temperature evolves with the inverse of the scale factor. Note that this does not require the particles to be able to interact with one another. The only thing acting on the particles in this case is the expansion of space.

Next, let's consider the case of a nonrelativistic decoupled particle species. Whereas the kinetic energy of a relativistic species scales with $E_k = p \propto a^{-1}$, the kinetic energy of a nonrelativistic particle evolves as $E_k = p^2/2m \propto a^{-2}$. Additionally, just as in the relativistic case, the number density of the particles falls with the increase of volume, $n \propto a^{-3}$. Combining these effects, one finds that the nonrelativistic particles retain an equilibrium distribution with a temperature that evolves as $T = (a_{\rm dc}/a)^2\,T_{\rm dc}$. Furthermore, in order to ensure that the number density continues to scale with $n \propto a^{-3}$, the particles must acquire a chemical potential, given by $\mu = m + (\mu_{\rm dc} - m)(a_{\rm dc}/a)^2$, where $\mu_{\rm dc}$ represents the chemical potential that was present at the time of decoupling (see Problem 4.10).

Although populations of decoupled particles that are either highly relativistic or nonrelativistic will retain an equilibrium distribution as the universe expands, this is not the case for particles with $T \sim m$. In this intermediate case, the distribution of particles that results will not match that of any equilibrium distribution.

4.7 The Boltzmann equation

As discussed in Sec. 4.4, thermal equilibrium is reached and maintained in most cosmological applications whenever the scattering rate for the relevant processes is much larger than the rate of Hubble expansion. In contrast, whenever the Hubble rate is much larger than the scattering rate, the particles in question will be decoupled from one another, and will not appreciably interact. Although this simple criterion for the demarcation between equilibrium and decoupling is adequate for a wide variety of applications in particle cosmology, we are sometimes interested in the detailed dynamics that play out as a collection of particles transitions away from, or into, a state of equilibrium. In these kinds of situations, one often employs variations of the Boltzmann equation, which describes the

statistical behavior of a thermodynamic system, regardless of whether or not it is in a state of equilibrium.

Although the most general form of the Boltzmann equation can be used to calculate the full evolution of the momentum distribution of a collection of particles, we're going to focus here on a simpler version that describes how the abundance of a given particle species evolves with time. Consider, for example, a collection of particle species that can undergo interactions of the form $1+2+\cdots \longleftrightarrow a+b+\cdots$, where here I am using numbers to denote particles on the left-hand side of the interaction, and letters to denote those on the right. The evolution of the number density of species 1 is governed by the following Boltzmann equation:

$$\frac{dn_1}{dt} = -3Hn_1 + \Gamma_{a+b+\cdots\rightarrow 1+2+\cdots} - \Gamma_{1+2+\cdots\rightarrow a+b+\cdots}. \tag{4.25}$$

The rightmost two terms represent the impact of interactions between particles, and are called the collision terms. The first of these collision terms is the production rate for the species in question, through the process $a+b+\cdots\rightarrow 1+2+\cdots$, while the second collision term corresponds to the destruction of such particles, via $1+2+\cdots\rightarrow a+b+\cdots$. The first term on the right-hand side accounts for the dilution resulting from the expansion of the universe. In the absence of any interactions, the solution to this equation is simply $n_1 \propto a^{-3}$. To summarize this equation, the evolution of the number density of a particle species is determined by the rate of Hubble expansion, along with the rates at which the species is produced and destroyed.

Conceptually, one can see that the collision terms should be the product of the number densities of the scattering particles, their scattering cross section, and their relative velocity. For processes of the type $1+2\rightarrow a+b$, for example, $\Gamma = n_1\, n_2\, \sigma\, v$. In practice, this is further complicated by factors such as the momentum dependance of $\sigma\, v$, and the quantum statistics of the final state particles. Integrating over the momenta of the various particles involved, we can write the 2-to-2 collision terms as

$$\begin{aligned}\Gamma_{1+2\rightarrow a+b} = \int &\left(\frac{dn_1}{d^3\vec{p}_1}d^3\vec{p}_1\right)\left(\frac{dn_2}{d^3\vec{p}_2}d^3\vec{p}_2\right)\left(\frac{S}{4\sqrt{(p_1\cdot p_2)^2-(m_1m_2)^2}}\right. \\ &\left.\times |\mathcal{M}_{1+2\rightarrow a+b}|^2(2\pi)^4\delta^4(p_1+p_2-p_a-p_b)\,\frac{1}{2E_a}\frac{d^3\vec{p}_a}{(2\pi)^3}\,\frac{1}{2E_b}\frac{d^3\vec{p}_b}{(2\pi)^3}\right) \\ &\times\left(\frac{\sqrt{(p_1\cdot p_2)^2-(m_1m_2)^2}}{E_1E_2}\right) \\ = \int &\, S\,\frac{g_1}{(2\pi)^3}\frac{d^3\vec{p}_1}{2E_1}f_1\,\frac{g_2}{(2\pi)^3}\frac{d^3\vec{p}_2}{2E_2}f_2\,\frac{1}{(2\pi)^3}\frac{d^3\vec{p}_a}{2E_a}\,\frac{1}{(2\pi)^3}\frac{d^3\vec{p}_b}{2E_b} \\ &\times |\mathcal{M}_{1+2\rightarrow a+b}|^2(2\pi)^4\delta^4(p_1+p_2-p_a-p_b),\end{aligned} \tag{4.26}$$

and similarly,

$$\begin{aligned}\Gamma_{a+b\rightarrow 1+2} = \int &\, S\,\frac{g_a}{(2\pi)^3}\frac{d^3\vec{p}_a}{2E_a}f_a\,\frac{g_b}{(2\pi)^3}\frac{d^3\vec{p}_b}{2E_b}f_b\,\frac{1}{(2\pi)^3}\frac{d^3\vec{p}_1}{2E_1}\,\frac{1}{(2\pi)^3}\frac{d^3\vec{p}_2}{2E_2} \\ &\times |\mathcal{M}_{a+b\rightarrow 1+2}|^2(2\pi)^4\delta^4(p_1+p_2-p_a-p_b),\end{aligned} \tag{4.27}$$

where p_i is the four-momentum, and $|\mathcal{M}|^2$ is the square of the spin-averaged amplitude for the process in question. As before, g_i is the number of internal degrees of freedom, and f_i is the phase space distribution of species i (see Eq. 4.1). Note that while $f_i = (e^{(E_i-\mu_i)/T_i} \pm 1)^{-1}$ for a species in kinetic equilibrium, these collision terms allow for the possibility that the particles in question are not described by an equilibrium distribution. S is equal to 1/2 if the two outgoing particles are identical, and is 1 otherwise.

There is a lot to unpack in these equations. In the first line of Eq. 4.26, the factors contained within large parentheses account, from left to right, for the number densities of species 1 and 2, the cross section for their annihilation (see Eq. 3.14), and the velocity between the incoming particles. While you might have thought that the velocity factor in these expressions would simply be given by the relative velocity between the incoming particles, $v_{\rm rel} = |\vec{v}_i - \vec{v}_j|$, this is not what was used in the equations above. To understand why, notice that the quantity on the left-hand side of the Boltzmann equation, dn/dt, is invariant under Lorentz transformations. After all, a Lorentz boost will contract the volume of a given region, reducing n, while increasing intervals in time by the same factor, resulting in no change in dn/dt. A similar argument applies to the $-3Hn$ term as well. From this, we can deduce that the collision term must also be Lorentz-invariant. But if I had used $v = v_{\rm rel}$ in this expression, it turns out that this would not have been the case.[2] In order for the scattering rate per volume to be the same in all frames of reference, we must instead employ what is sometimes called the Møller velocity, $v_{\rm Moll}$, which is defined as follows [21]:

$$v_{\rm Moll} \equiv \frac{\sqrt{(p_i \cdot p_j)^2 - (m_i m_j)^2}}{E_i E_j}. \tag{4.28}$$

Notice that the Møller velocity is equal to the velocity of particle i in the rest frame of particle j, and thus $v_{\rm Moll} \approx v_{\rm rel}$ in the nonrelativistic limit (see Problem 4.11).

In situations in which $f \ll 1$, Eqs. 4.26 and 4.27 will accurately describe the collision terms for 2-to-2 processes. If this is not the case, however, we will need to include a factor of $1 \pm f_i$ for each outgoing particle, with the plus sign being appropriate for outgoing bosons, and the minus sign for outgoing fermions. These quantum statistical factors account for the phenomena of Bose enhancement and Pauli blocking. For example, these suppress the production of fermions within those regions of phase space that are already significantly filled. These factors are generally not important for particle species that are nonrelativistic, or more generally when $T \ll m - \mu$. In this case, we can ignore the effects of Pauli blocking and Bose enhancement, substituting $(1 \pm f) \to 1$.

Incorporating these quantum statistics factors, and generalizing to the case of processes involving any number of particles, we can write the collision terms of the Boltzmann equation as follows:

$$\begin{aligned}\Gamma_{1+2+\cdots \to a+b+\cdots} = \int S \prod_{i=1}^{i_{\rm max}} \frac{g_i}{(2\pi)^3} \frac{d^3\vec{p}_i}{2E_i} f_i \prod_{j=a}^{j_{\rm max}} \frac{1}{(2\pi)^3} \frac{d^3\vec{p}_j}{2E_j} (1 \pm f_j) \\ \times |\mathcal{M}_{1+2+\cdots \to a+b+\cdots}|^2 \, (2\pi)^4 \, \delta^4\left(\sum_{i=1}^{i_{\rm max}} p_i - \sum_{j=a}^{j_{\rm max}} p_j \right)\end{aligned} \tag{4.29}$$

[2] Note that σ is defined as the cross section in the target's rest frame, so it is the quantity $v\, n_i\, n_j$ that must be Lorentz-invariant.

and

$$\Gamma_{a+b+\cdots\rightarrow 1+2+\cdots} = \int S \prod_{j=a}^{j_{\rm max}} \frac{g_j}{(2\pi)^3} \frac{d^3\vec{p}_j}{2E_j} f_j \prod_{i=1}^{i_{\rm max}} \frac{1}{(2\pi)^3} \frac{d^3\vec{p}_i}{2E_i} (1 \pm f_i) \tag{4.30}$$

$$\times |\mathcal{M}_{a+b+\cdots\rightarrow 1+2+\cdots}|^2 \, (2\pi)^4 \, \delta^4\Bigg(\sum_{i=1}^{i_{\rm max}} p_i - \sum_{j=a}^{j_{\rm max}} p_j \Bigg).$$

In each of these expressions, S is equal to the product of $1/s!$ for each group of identical outgoing particles.

The nature of the interactions under consideration is entirely encoded within the squared amplitudes, $|\mathcal{M}_{a+b+\cdots\rightarrow 1+2+\cdots}|^2$ and $|\mathcal{M}_{1+2+\cdots\rightarrow a+b+\cdots}|^2$, which are determined by the couplings, spins, masses, and other characteristics of the particle species in question. Up to this point, I have written these two squared amplitudes as if they were distinct and independent of one another. In practice, however, we will often find ourselves dealing with reversible processes, in which case these two squared amplitudes will be necessarily equal, $|\mathcal{M}_{a+b+\cdots\rightarrow 1+2+\cdots}|^2 = |\mathcal{M}_{1+2+\cdots\rightarrow a+b+\cdots}|^2$. In light of the CPT theorem, the condition of reversibility is equivalent to assuming that the processes in question are symmetric with respect to the combined transformations of charge and parity, CP.[3] Unless we are considering special cases in which CP symmetry is significantly violated, we will typically assume that the squared amplitude appearing in the production term of the Boltzmann equation is the same as the squared amplitude in the destruction term.

The version of the Boltzmann equation described above includes only one class of interactions: those of the form $1 + 2 + \cdots \longleftrightarrow a + b + \cdots$. One could imagine, however, many applications in which a given particle species could be produced or destroyed through a number of different interactions. If I wanted to write down the complete Boltzmann equation to calculate the evolution of the number density of electrons at $T \sim 100$ MeV, for example, I would need to include collision terms corresponding to $e^+e^- \leftrightarrow \gamma\gamma$, $e^+e^- \leftrightarrow \mu^+\mu^-$, $e^+e^- \leftrightarrow \pi\pi$, and $e^+e^- \leftrightarrow \nu\bar{\nu}$. In addition, I would have to keep track of the abundances of each of the particle species that can interact to produce electrons, accounting for processes such as $\mu^+\mu^- \leftrightarrow \gamma\gamma$, for example. This would lead to a situation in which I would have to solve a complex system of coupled Boltzmann equations in order to calculate the evolution of the abundances of all of the particle species in question. Fortunately, in many situations, one can substantially reduce the complexity of such problems by taking advantage of the fact that many of the species in question will closely follow equilibrium distributions.

A particle species will often remain in kinetic equilibrium well after it has departed from chemical equilibrium, in which case the momentum dependance of f_i will be the same as that of $f_i^{\rm Eq}$, differing only in their normalization. With this in mind, it is useful to consider the cross section for a given process, averaged over the momenta of an equilibrium distribution. For a 2-to-2 process, this can be written as

$$\langle \sigma v \rangle \equiv \frac{\int dn_1^{\rm Eq} dn_2^{\rm Eq} \, (\sigma v_{\rm Moll})_{1+2\rightarrow a+b}}{\int dn_1^{\rm Eq} dn_2^{\rm Eq}} \tag{4.31}$$

[3] The CPT theorem states that the combined symmetries of charge, parity, and time are held in any Lorentz-invariant, local quantum field theory with a Hermitian Hamiltonian.

$$= \frac{\int d^3\vec{p}_1 d^3\vec{p}_2 f_1^{\rm Eq} f_2^{\rm Eq} \, (\sigma \, v_{\rm Moll})_{1+2\to a+b}}{\int d^3\vec{p}_1 d^3\vec{p}_2 f_1^{\rm Eq} f_2^{\rm Eq}}.$$

Note that while $\sigma\, v$ is generally a function of the velocity between the incoming particles, the quantity $\langle \sigma\, v \rangle$ is a function of the temperature of the particles in question. Furthermore, the principle of detailed balance states that, in equilibrium, the rate for a process and its reverse must be equal, and thus

$$\int f_1^{\rm Eq} f_2^{\rm Eq} (\sigma\, v_{\rm Moll})_{1+2\to a+b} \, d^3\vec{p}_1 d^3\vec{p}_2 = \int f_a^{\rm Eq} f_b^{\rm Eq} (\sigma\, v_{\rm Moll})_{a+b\to 1+2} \, d^3\vec{p}_a d^3\vec{p}_b. \tag{4.32}$$

This allows us to relate a given cross section to that of its inverse process.

We are now in a position to write the Boltzmann equation in terms of the thermally averaged cross section, and the number densities of the particle species involved:

$$\begin{aligned}
\frac{dn_1}{dt} &= -3Hn_1 + n_a n_b \, \langle \sigma\, v \rangle_{a+b\to 1+2} - n_1 n_2 \, \langle \sigma\, v \rangle_{1+2\to a+b} \qquad (4.33)\\
&= -3Hn_1 + n_a n_b \, \frac{\int f_a^{\rm Eq} f_b^{\rm Eq} (\sigma\, v)_{a+b\to 1+2} \, d^3\vec{p}_a d^3\vec{p}_b}{\int f_a^{\rm Eq} f_b^{\rm Eq} \, d^3\vec{p}_a d^3\vec{p}_b} - n_1 n_2 \, \langle \sigma\, v \rangle_{1+2\to a+b} \\
&= -3Hn_1 + n_a n_b \, \frac{\int f_1^{\rm Eq} f_2^{\rm Eq} (\sigma\, v)_{1+2\to a+b} \, d^3\vec{p}_1 d^3\vec{p}_2}{\int f_a^{\rm Eq} f_b^{\rm Eq} \, d^3\vec{p}_a d^3\vec{p}_b} - n_1 n_2 \, \langle \sigma\, v \rangle_{1+2\to a+b} \\
&= -3Hn_1 + n_a n_b \, \langle \sigma\, v \rangle_{1+2\to a+b} \frac{\int f_1^{\rm Eq} f_2^{\rm Eq} \, d^3\vec{p}_1 d^3\vec{p}_2}{\int f_a^{\rm Eq} f_b^{\rm Eq} \, d^3\vec{p}_a d^3\vec{p}_b} - n_1 n_2 \, \langle \sigma\, v \rangle_{1+2\to a+b} \\
&= -3Hn_1 + \langle \sigma\, v \rangle_{1+2\to a+b} \left[n_a n_b \left(\frac{n_1^{\rm Eq} \, n_2^{\rm Eq}}{n_a^{\rm Eq} n_b^{\rm Eq}} \right) - n_1 n_2 \right].
\end{aligned}$$

Variants of this Boltzmann equation are used in a wide variety of applications in particle cosmology. This includes the decoupling of photons from electrons at the time of recombination, several of the processes involved in Big Bang nucleosynthesis, and the thermal freeze out of various dark matter candidates. There are also examples of decoupling that do not require the use of the Boltzmann equation. In our discussion of neutrino decoupling in Sec. 4.5, for example, the Boltzmann equation was not necessary. This is because the kinetic and chemical decoupling of the neutrinos occurred when all of the species in question were highly relativistic, and at a time when little else other than the expansion of space was occurring.

At this point, it may be helpful to look at a couple of specific applications of the Boltzmann equation, which we will explore in greater detail in later chapters of this book. First, let's consider the evolution of the number density of free electrons during the era of recombination, when free electrons and protons were combining to form neutral hydrogen

atoms. In this case, we can apply the following Boltzmann equation:

$$\frac{dn_e}{dt} = -3Hn_e + n_e^{\rm Eq} n_p^{\rm Eq} \langle \sigma\, v \rangle \left[\left(\frac{n_H\, n_\gamma}{n_H^{\rm Eq} n_\gamma^{\rm Eq}} \right) - \left(\frac{n_e\, n_p}{n_e^{\rm Eq} n_p^{\rm Eq}} \right) \right] \tag{4.34}$$

$$= -3Hn_e + \langle \sigma\, v \rangle \left[\left(\frac{n_H\, n_e^{\rm Eq} n_p^{\rm Eq}}{n_H^{\rm Eq}} \right) - n_e^2 \right],$$

where $\langle \sigma\, v \rangle$ is the thermally averaged cross section for the process $e^- + p \to H + \gamma$. In the second line, I have made use of the fact that the overall charge neutrality of the universe requires $n_e = n_p$, and that the number density of photons was very close to its equilibrium value during this period of time. To make further use of this expression, we will need to determine the equilibrium value of $(n_e n_p / n_H)$, as described in Chapter 6.

Another relatively straightforward application of the Boltzmann equation is to calculate the evolution and thermal freeze out of a dark matter candidate in the form of a weakly interacting massive particle (WIMP). Let's posit that these particles, X, can annihilate into a pair of lighter Standard Model particles, $XX \to SS'$. In this case, the evolution of the number density of the WIMPs is given by

$$\frac{dn_X}{dt} = -3Hn_X + (n_X^{\rm Eq})^2 \langle \sigma\, v \rangle \left[\frac{n_S n_{S'}}{n_S^{\rm Eq} n_{S'}^{\rm Eq}} - \frac{n_X^2}{(n_X^{\rm Eq})^2} \right] \tag{4.35}$$

$$\approx -3Hn_X + \langle \sigma\, v \rangle \left[(n_X^{\rm Eq})^2 - n_X^2 \right],$$

where the thermally averaged cross section in this expression corresponds to the process of WIMP annihilation. In the last line, I have (reasonably) assumed that the Standard Model particles are present in approximately their equilibrium abundance with no significant chemical potential. At early times, when $\langle \sigma\, v \rangle n_X^{\rm Eq} \gg 3H$, the X abundance was kept very close to its equilibrium value. As the universe expanded and the temperature dropped, however, the rate of WIMP annihilation fell below that of Hubble expansion, causing the comoving number density of WIMPs to freeze out and become approximately constant, $n_X \propto a^{-3}$. We will discuss the thermal freeze out of dark matter in more detail in Chapter 9.

Before ending this chapter, I'll mention that the Boltzmann equation can also be written in terms of the ratio of the number density of a particle species to the total entropy density, $Y_i \equiv n_i / s$. For example, in terms of Y_X, I can rewrite Eq. 4.35 as follows:

$$\frac{dY_X}{dt} = s \langle \sigma\, v \rangle \left[(Y_X^{\rm Eq})^2 - Y_X^2 \right]. \tag{4.36}$$

Notice that, in this form, the effects of Hubble expansion are not reflected as an independent term, but are instead built into dY_X/dt. Dividing both sides by aH, we can write this equation in terms of the scale factor:

$$\frac{dY_X}{da} = \frac{s \langle \sigma\, v \rangle}{aH} \left[(Y_X^{\rm Eq})^2 - Y_X^2 \right]. \tag{4.37}$$

Since $\langle \sigma v \rangle$ is a function of temperature, it is often convenient to write this in terms of the quantity $x \equiv m/T$:

$$\frac{dY_X}{dx} = -\frac{1}{3H}\frac{ds}{dx}\langle \sigma v \rangle \left[(Y_X^{\rm Eq})^2 - Y_X^2 \right]. \tag{4.38}$$

Additional reading

I learned a lot of what I know about thermodynamics and statistical mechanics in a course taught by Randy Durand, which used Landau and Lifshiz's *Statistical Physics*. For books that explore this topic within the context of particle cosmology, I strongly recommend Dodelson's *Modern Cosmology*, and Kolb and Turner's *The Early Universe*.

Problems

4.1 Numerically integrate Eq. 4.5 to obtain the number density, energy density, and pressure of a particle species in equilibrium. Show that your code recovers the results given in Eqs. 4.6 and 4.8 in the appropriate limits.

4.2 Use Eq. 4.10 to calculate the value of $g_\star$ for the Standard Model particle content, evaluated at temperatures of 1 MeV, 10 MeV, 20 GeV, 100 GeV, and 1 TeV. Use the approximation that all particles with $m < T$ are relativistic, and neglect all particle species with $m > T$.

4.3 Repeat the previous problem without making the same approximations. Instead, numerically integrate the energy densities using Eq. 4.5 to obtain a more accurate value for $g_\star$ at each temperature. Compare these results to those obtained in the previous problem. At which temperatures are the approximations that were used in that problem least valid? Why?

4.4 Let the scale factor of the universe be a_i when $T = 1$ TeV. Use entropy conservation to determine the temperature when $a = 10 \times a_i$, $10^3 \times a_i$, and $10^6 \times a_i$.

4.5 Imagine a universe in which only photons, electrons, and muons (and their antiparticles) exist. How does the temperature of the thermal bath evolve with scale factor? How would this relationship have been different if muons had not existed?

4.6 Consider the following processes: $e^+e^- \leftrightarrow \mu^+\mu^-$, $\mu^+\mu^- \leftrightarrow \gamma\gamma$, and $\gamma\gamma \leftrightarrow \gamma\gamma\gamma$. From the fact that these processes are possible, what relationships can you conclude about these particles' chemical potentials (when in chemical equilibrium)?

4.7 Consider a particle species X with a mass of $m_X = 100$ GeV and two internal degrees of freedom, which can annihilate into Standard Model fermions, $XX \rightarrow f\bar{f}$, with a thermally averaged cross section of $\langle \sigma v \rangle = 10^{-26}\,{\rm cm}^3/{\rm s}$. Assuming that the X population is in equilibrium with the Standard Model bath, calculate the temperature at which its annihilation rate equals that of the Hubble rate.

4.8 Imagine that instead of at a few MeV, neutrino decoupling had taken place at 200 MeV, above the masses of the muon and pions. Using entropy conservation, calculate the value of T_ν/T at $T \ll m_e$.

4.9 From the temperature of the cosmic neutrino background, $T_\nu = 1.945$ K, calculate the number density and energy density of these particles. Treat the neutrinos as massless.

4.10 Consider a nonrelativistic species that decoupled with an equilibrium distribution and zero chemical potential. As described in Sec. 4.6, these particles retain

their equilibrium distribution as the universe expands, evolving as $T \propto a^{-2}$ and $\mu = m[1 - (a_{\rm dc}/a)^2]$. Show that the comoving number density of these particles remains constant as this evolution proceeds.

4.11 Consider two particles. Show that the Møller velocity (Eq. 4.28) is equal to the relative velocity between the two particles in the rest frame of either particle.

4.12 Show that the Boltzmann equation given in Eq. 4.36 is equivalent to Eq. 4.35.

5

The Interactions of High-Energy Particles in Astrophysics

In this chapter, I shift away from the early universe and instead focus on the high-energy interactions that take place in our universe today. In particular, I review some of the most important interactions that are experienced by high-energy electrons, nucleons, nuclei, and photons as they travel through our galaxy and universe. This information will be important in our later discussions of cosmic rays, gamma-ray astronomy, neutrino astronomy, and indirect searches for dark matter, as discussed in Chapters 15, 16, and 19.

5.1 Interactions of high-energy electrons

High-energy electrons interact with gas, magnetic fields, and radiation fields through the processes of bremsstrahlung, synchrotron, and inverse Compton scattering, respectively. In this section, we'll explore the energy losses that result from these interactions, as well as the gamma-ray, X-ray, and radio signals that they can generate.

5.1.1 Bremsstrahlung

As energetic charged particles pass through electric fields, they can radiate photons and lose energy. This process, which is known as bremsstrahlung, is particularly important for electrons,[1] causing them to lose significant amounts of energy through the emission of X-rays or gamma rays [22].

A collection of electrons in an energy distribution, dN_e/dE_e, will produce a spectrum of bremsstrahlung radiation that is given by the following:

$$\frac{dN_\gamma}{dE_\gamma} = \sum_i n_i \int 2 \frac{dN_e}{dE_e} \frac{d\sigma_i}{dE_\gamma} \, dE_e, \tag{5.1}$$

[1] Throughout much of this chapter, I will use the word "electron" to refer to both electrons and positrons. After all, positrons experience bremsstrahlung, synchrotron, and inverse Compton processes in the same way as electrons.

where n_i is the number density of gas particles of species i, and $d\sigma_i/dE_\gamma$ is the differential cross section for bremsstrahlung on that target species. This cross section is given by [23]

$$\frac{d\sigma_i}{dE_\gamma} = \frac{3\alpha\sigma_t}{8\pi E_\gamma}\left\{\left[1+\left(1-\frac{E_\gamma}{E_e}\right)^2\right]\phi_1^i - \frac{2}{3}\left(1-\frac{E_\gamma}{E_e}\right)\phi_2^i\right\}, \tag{5.2}$$

where α is the fine structure constant, σ_t is the Thomson cross section, and $\phi_{1,2}^i$ are functions that depend on the characteristics of the target. For the case of fully ionized gas, these functions are given by

$$\phi_1^{\rm ion} = \phi_2^{\rm ion} = 4(Z^2+Z)\left\{\ln\left[\frac{2E_e}{m_e}\left(\frac{E_e-E_\gamma}{E_\gamma}\right)\right] - \frac{1}{2}\right\}, \tag{5.3}$$

where Z is the atomic number of the gas. In the $E_e \gg m_e$ limit, these functions for neutral atoms are given by $\phi_1^{\rm H} \approx 45.79$, $\phi_2^{\rm H} \approx 44.46$, $\phi_1^{\rm He} \approx 134.60$, $\phi_2^{\rm He} \approx 131.40$, $\phi_1^{\rm H_2} \approx 2\phi_1^{\rm H}$, and $\phi_2^{\rm H_2} \approx 2\phi_2^{\rm H}$.

From the differential cross section and the density of targets, we can determine the rate at which relativistic electrons lose energy through the process of bremsstrahlung emission:

$$\frac{dE_e}{dt} = -\sum_i n_i \int_0^{E_e} \frac{d\sigma_i}{dE_\gamma} E_\gamma\, dE_\gamma. \tag{5.4}$$

In the case of a target consisting of neutral gas, ϕ_1 and ϕ_2 are independent of energy, allowing us to write the energy loss rate as follows:

$$\begin{aligned}\frac{dE_e}{dt} &= -\sum_i n_i \int_0^{E_e} \frac{3\alpha\sigma_t}{8\pi}\left\{\left[1+\left(1-\frac{E_\gamma}{E_e}\right)^2\right]\phi_1^i - \frac{2}{3}\left(1-\frac{E_\gamma}{E_e}\right)\phi_2^i\right\} dE_\gamma \\ &= -\sum_i \frac{\alpha\sigma_t E_e n_i}{8\pi}\left[4\phi_1^i - \phi_2^i\right].\end{aligned} \tag{5.5}$$

From this, we can derive the timescale for bremsstrahlung losses. For the case of scattering with neutral gas, this is given by

$$\begin{aligned}\tau_{\rm Brem} &\equiv \left(\frac{1}{E_e}\frac{dE_e}{dt}\right)^{-1} \\ &= \frac{8\pi}{\alpha\sigma_t \sum_i n_i\left[4\phi_1^i - \phi_2^i\right]} \\ &\approx 1.2\times 10^{15}\ {\rm s} \times \left(\frac{1\ {\rm cm}^{-3}}{n}\right)\left(\frac{138.7}{4\phi_1-\phi_2}\right).\end{aligned} \tag{5.6}$$

While bremsstrahlung also occurs for protons and other charged particles, the timescale for such losses is proportional to the square of the mass of the radiating particle (since $\sigma_t \propto m_e^{-2}$). As a result, the process of bremsstrahlung can usually be ignored when considering protons or other heavy particles.

Note that the timescale for bremsstrahlung energy losses is independent of the electron's energy. The spectrum of the photons that are produced through this mechanism exhibits a shape that approximately mirrors that of the electrons themselves, and extends up to the energy of the radiating electron. This process often provides the dominant cooling mechanism for cosmic-ray electrons in the ∼MeV-GeV energy range. At lower energies, ionization losses are more important, while at higher energies synchrotron and inverse Compton scattering are typically more significant.

5.1.2 Synchrotron

In the presence of magnetic fields, ultra-relativistic electrons gyrate and produce electromagnetic radiation. The total energy loss rate resulting from this process is given by [23]

$$\frac{dE_e}{dt} = -2\sigma_t \rho_{\rm mag} v_e^2 \gamma_e^2 \ \sin^2 \alpha_p, \tag{5.7}$$

where σ_t is again the Thomson cross section, v_e and γ_e are the velocity and the Lorentz factor of the electron, and $\rho_{\rm mag}$ is the energy density of the magnetic field. The quantity α_p is the angle between the electron's trajectory and the orientation of the magnetic field, known as the pitch angle. Averaging over the pitch angle, $\langle \sin^2 \alpha_p \rangle = \int_0^\pi \sin^3 \alpha_p d\alpha_p = 2/3$, the energy loss rate becomes

$$\begin{aligned} \frac{dE_e}{dt} &= -\frac{4}{3}\sigma_t \rho_{\rm mag} v_e^2 \gamma_e^2 \\ &\approx -2.28 \times 10^{-17}\ {\rm GeV/s} \times \left(\frac{E_e}{\rm GeV}\right)^2 \left(\frac{B}{3\,\mu{\rm G}}\right)^2. \end{aligned} \tag{5.8}$$

In the last line of this expression, I have adopted $v_e \approx c$ and related the energy density in the magnetic field to the strength of that field:

$$\begin{aligned} \rho_{\rm mag} &= \frac{B^2}{2\mu_0} \\ &\approx 0.224\ {\rm eV/cm^3} \times \left(\frac{B}{3\,\mu{\rm G}}\right)^2. \end{aligned} \tag{5.9}$$

In the presence of μG-scale magnetic fields and gas densities on the order of $\sim 1\,{\rm cm}^{-3}$, such as are typically found in the interstellar medium of the Milky Way, synchrotron energy losses are generally larger than those from bremsstrahlung for $E_e \gtrsim \mathcal{O}(10\,{\rm GeV})$.

The intensity of the synchrotron radiation from a relativistic electron peaks near what is known as the critical frequency, which is given by

$$\begin{aligned} \nu_c &= \frac{3}{2}\gamma_e^2 \nu_g \sin\alpha_p \\ &\approx 4.5 \times 10^7\ {\rm Hz} \times \left(\frac{E_e}{\rm GeV}\right)^2 \left(\frac{B}{3\,\mu{\rm G}}\right) \sin\alpha_p, \end{aligned} \tag{5.10}$$

where $\nu_g = eB/2\pi m_e \approx 8.4\,{\rm Hz} \times (B/3\mu{\rm G})$ is the the nonrelativistic gyrofrequency. The total power per unit frequency emitted as synchrotron radiation from a relativistic electron (also

known as the emissivity, j) is given by

$$j(\nu) \equiv 2\pi\,\nu\,\frac{dN_\gamma}{d\nu} \tag{5.11}$$
$$= \frac{\sqrt{3}\,\alpha^{3/2} B \sin\alpha_p}{\sqrt{\pi}\,m_e}\,F(\nu/\nu_c),$$

where

$$F(x) \equiv x\int_x^\infty K_{5/3}(z)dz, \tag{5.12}$$

and where $F_{5/3}$ is the modified Bessel function of order 5/3. In the limits of small or large x, $F(x)$ can be approximated by

$$F(x) = \begin{cases} \dfrac{4\pi}{\sqrt{3}\,\Gamma(\frac{1}{3})}\left(\dfrac{x}{2}\right)^{1/3}, & x \ll 1, \\ \\ \left(\dfrac{\pi}{2}\right)^{1/2} x^{1/2}\,e^{-x}, & x \gg 1, \end{cases} \tag{5.13}$$

where $\Gamma(\frac{1}{3}) \approx 2.67$ is the gamma function evaluated at 1/3. Combining this with Eq. 5.11, the synchrotron spectrum from a single electron behaves as $j(\nu) \propto \nu^{1/3}$ at frequencies well below the critical frequency, and as $j(\nu) \propto \nu^{1/2}\,e^{-\nu/\nu_c}$ at high frequencies.

To estimate the spectrum of the synchrotron radiation emitted from a population of electrons, let's adopt the simplifying approximation that each electron radiates the entirety of its power at a frequency equal to its critical frequency. While this is not exactly realistic, the power does climb until $\nu \sim \nu_c$ and fall exponentially at higher frequencies, leading to results that are not very different from those that would be obtained in a more careful calculation. Under this approximation, the total synchrotron spectrum can be written as

$$j(\nu) = \left(\frac{-dE_e}{dt}\right)\left(\frac{dN_e}{dE_e}\right)\left(\frac{\Delta E_e}{\Delta\nu_c}\right), \tag{5.14}$$

where $-dE_e/dt$ is the synchtron energy loss rate as given in Eq. 5.8, dN_e/dE_e is the spectrum of the radiating electrons, and $\Delta\nu_c$ is the degree to which the critical frequency of the synchrotron emission changes as an electron cools from E_e to $E_e - \Delta E_e$:

$$\begin{aligned} \Delta\nu_c &= \nu_c(E_e) - \nu_c(E_e - \Delta E_e) \\ &= \frac{3}{2}\left(\frac{E_e}{m_e}\right)^2 \frac{eB\sin\alpha_p}{2\pi\,m_e} - \frac{3}{2}\left(\frac{E_e - \Delta E_e}{m_e}\right)^2 \frac{eB\sin\alpha_p}{2\pi\,m_e} \\ &\approx \frac{3E_e\Delta E_e eB\sin\alpha_p}{2\pi\,m_e^3} \\ &= \frac{2\nu_c\Delta E_e}{E_e}. \end{aligned} \tag{5.15}$$

For the case of a power-law spectrum of electrons, $dN_e/dE_e = AE_e^{-p}$, the spectrum of synchrotron emission is approximately given by

$$j(\nu) = \left(\frac{4\sigma_t}{3}\frac{B^2}{2\mu_0}\frac{E_e^2}{m_e^2}\right)\left(AE_e^{-p}\right)\left(\frac{E_e}{2\nu}\right) \tag{5.16}$$
$$\approx \frac{A\sigma_t B^2}{3\mu_0 m_e^2 \nu}\left(\frac{2\pi m_e^3 \nu}{eB}\right)^{(3-p)/2}.$$

where in the second line I have made the substitution $E_e \approx (2\pi m_e^3 \nu/eB)^{1/2}$. Whereas the spectrum of bremsstrahlung emission takes on approximately the same spectral shape as the electrons that produce that radiation, we can see from this exercise that a population of electrons with a power-law index p will produce a spectrum of synchrotron radiation that takes the form of $dN_\gamma/d\nu \propto j(\nu)/\nu \propto \nu^{-(p+1)/2}$.

Observations of diffuse radio emission from the Milky Way indicate that cosmic-ray electrons are producing synchrotron emission with a spectral slope that varies from $j(\nu) \propto \nu^{-0.4}$ at $\nu \sim 10$ MHz to $j(\nu) \propto \nu^{-0.8}$ at GHz frequencies and above. Using Eq. 5.16, this implies that the Milky Way's cosmic-ray electrons have a spectral index that is quite hard at low frequencies, $p \sim 1.8$, but that softens to $p \sim 2.6$ at GeV energies and above.

5.1.3 Inverse compton scattering

Over approximately the same range of energies over which synchrotron plays an important role, electrons can lose energy by scattering with low-energy photons. This process, known as inverse Compton scattering, produces gamma rays by transferring a portion of the cosmic-ray electron's energy to the target photon.

In the limit of $E_e \ll m_e^2/\epsilon_\gamma$, where ϵ_γ is the energy of a target photon, electrons lose energy through inverse Compton scattering at the following rate [23]:

$$\frac{dE_e}{dt} = -\frac{4}{3}\sigma_t \rho_{\rm rad} v_e^2 \gamma_e^2 \tag{5.17}$$
$$\approx -1.02 \times 10^{-16}\,{\rm GeV/s} \times \left(\frac{E_e}{\rm GeV}\right)^2 \left(\frac{\rho_{\rm rad}}{\rm eV/cm^3}\right),$$

where $\rho_{\rm rad}$ is the energy density of the target photons. In this limit, the energy loss rate from inverse Compton scattering takes on an identical form to that from synchrotron (see Eq. 5.8).

An electron with $E_e \ll m_e^2/\epsilon_\gamma$, traveling through a uniform distribution of photons of energy ϵ_γ, will produce the following spectrum of inverse Compton photons per unit time:

$$\frac{dN_\gamma}{dE_\gamma}(E_\gamma, E_e) \approx \frac{3\pi\sigma_t n_\gamma m_e^4 E_\gamma}{8E_e^4\epsilon_\gamma^2}\left[2E_\gamma \ln\left(\frac{E_\gamma m_e^2}{4E_e^2\epsilon_\gamma}\right) + E_\gamma + \frac{4E_e^2\epsilon_\gamma}{m_e^2} - \frac{E_\gamma^2 m_e^2}{2E_e^2\epsilon_\gamma}\right]. \tag{5.18}$$

Comparing the energy loss rate given in Eq. 5.17 to the scattering rate of $\sigma_t\rho_{\rm rad}/\epsilon_\gamma$, we can determine the mean energy of a photon produced through inverse Compton scattering:

$$\langle E_\gamma \rangle = \frac{4\sigma_t \rho_{\rm rad} v_e^2 \gamma_e^2/3}{\sigma_t \rho_{\rm rad}/\epsilon_\gamma} \tag{5.19}$$

$$= \frac{4}{3} v_e^2 \gamma_e^2 \epsilon_\gamma$$

$$\approx 5\,\mathrm{MeV} \times \left(\frac{E_e}{\mathrm{GeV}}\right)^2 \left(\frac{\epsilon_\gamma}{\mathrm{eV}}\right).$$

The maximum energy that a photon can have after undergoing inverse Compton scattering is given by $E_\gamma^{\mathrm{max}} = 4E_e^2 \epsilon_\gamma / m_e^2$, corresponding to the case in which the electron and target photon collide head-on and the photon then emerges from the collision traveling in the opposite direction. By comparing this maximum value of the photon energy to the average value given above, we see that inverse Compton scattering in the $E_e \ll m_e^2/\epsilon_\gamma$ limit generally produces photons with a spectrum that is fairly sharply peaked near the maximum energy allowed by the kinematics of the problem.

Notice that for $E_e > 3m_e^2/4\epsilon_\gamma$, Eq. 5.19 indicates that the average energy of a scattered photon will be greater than that of the incoming electron, clearly violating energy conservation. The resolution to this inconsistency is that so far we have only considered inverse Compton scattering in the so-called Thomson limit, in which $E_e \ll m_e^2/\epsilon_\gamma$. For higher energy electrons, we must replace the Thomson scattering cross section in the above expressions with the Klein-Nishina cross section. Considering a spectrum of target radiation that is described by a blackbody of temperature T, this leads to the following energy loss rate:

$$\frac{dE_e}{dt} = -\frac{4}{3}\sigma_t \gamma_e^2 \int \frac{dn_\gamma}{d\epsilon_\gamma}\, \epsilon_\gamma \left(\frac{\gamma_k^2}{\gamma_k^2 + \gamma_e^2}\right) d\epsilon_\gamma, \tag{5.20}$$

where $\gamma_k = 3\sqrt{5} m_e / 8\pi T$. For $E_e \ll m_e^2/T$, the term in brackets approaches unity and we recover the energy loss rate found in the Thomson limit. At higher energies, however, the energy loss rate is suppressed by an additional factor of $\sim 45 m_e^4 / 64\pi^2 T^2 E_e^2$ relative to that in the Thomson regime.

Cosmic-ray electrons in the local interstellar medium encounter radiation in the form of starlight (with $\rho_{\mathrm{star}} \sim 0.6\,\mathrm{eV/cm^3}$, $T_{\mathrm{star}} \sim 5{,}000\,\mathrm{K} \sim 0.4\,\mathrm{eV}$), as well as significant infrared radiation ($\rho_{\mathrm{IR}} \sim 0.6\,\mathrm{eV/cm^3}$, $T_{\mathrm{IR}} \sim 20\,\mathrm{K} \sim 0.0017\,\mathrm{eV}$), and of course the CMB ($\rho_{\mathrm{CMB}} \approx 0.26\,\mathrm{eV/cm^3}$, $T_{\mathrm{CMB}} \approx 2.7\,\mathrm{K} \approx 0.00023\,\mathrm{eV}$). At low energies, the energy losses from inverse Compton scattering are dominated by interactions with optical and infrared photons, due to their larger energy densities. For more energetic electrons, however, Klein-Nishina suppression lessens the impact of this higher frequency radiation, causing lower energy target photons to play a more important role.

Including the full Klein-Nishina cross section, the spectrum of inverse Compton photons produced by an electron of energy, E_e, is given by

$$\frac{dN_\gamma}{dE_\gamma}(E_\gamma, E_e) \propto \int \frac{dn_\gamma}{d\epsilon_\gamma}(\epsilon_\gamma)\, \frac{d\sigma_{\mathrm{ICS}}}{dE_\gamma}(E_\gamma, \epsilon_\gamma, E_e)\, d\epsilon_\gamma, \tag{5.21}$$

where $dn_\gamma/d\epsilon_\gamma$ is the spectrum of target photons, and the normalization is set such that the energy lost by the electron according to Eq. 5.20 is equal to the integral $\int dE_\gamma\, E_\gamma\, dN_\gamma/dE_\gamma$. The inverse Compton scattering cross section in the above expression is given by the following:

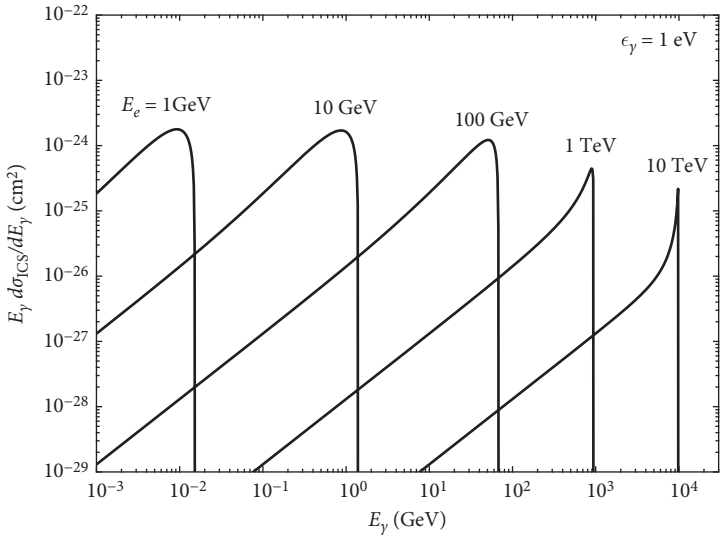


Figure 5.1 The differential cross section for the inverse Compton scattering of a high-energy electron with a 1 eV target photon, for several values of E_e. At very high energies ($E_e \gg m_e^2/\epsilon_\gamma$), this process takes place in the Klein-Nishina regime, leading to a smaller overall cross section that peaks sharply at $E_\gamma \approx E_e$.

$$\frac{d\sigma_{\rm ICS}}{dE_\gamma}(E_\gamma, \epsilon_\gamma, E_e) = \frac{3\sigma_t m_e^2}{4\epsilon_\gamma E_e^2}\left[1 + \left(\frac{z^2}{2(1-z)}\right) + \left(\frac{z}{\beta(1-z)}\right) - \left(\frac{2z^2}{\beta^2(1-z)}\right) - \left(\frac{z^3}{2\beta(1-z)^2}\right) + \left(\frac{2z}{\beta(1-z)}\right)\ln\left(\frac{\beta(1-z)}{z}\right)\right], \tag{5.22}$$

where $z \equiv E_\gamma/E_e$ and $\beta \equiv 4\epsilon_\gamma E_e/m_e^2$. In the Klein-Nishina regime, the spectrum of inverse Compton emission peaks strongly at energies only slightly below E_e. This is illustrated in Figure 5.1, were I have plotted the differential cross section for inverse Compton scattering for $\epsilon_\gamma = 1\,{\rm eV}$ and for several values of E_e. For $E_e \ll m_e^2/\epsilon_\gamma \approx 250\,{\rm GeV}$, we are in the Thomson regime and $d\sigma_{\rm ICS}/dE_\gamma$ is approximately independent of E_γ up to a maximum energy given by $E_\gamma = 4\epsilon_\gamma E_e^2/m_e^2$. At higher energies, this process takes place in the Klein-Nishina regime, leading to a smaller overall cross section that peaks increasingly sharply at $E_\gamma \approx E_e$.

5.2 Interactions of high-energy nucleons and nuclei

So far, the interactions described in this chapter have been the result of the electromagnetic force. Nucleons and nuclei, however, also interact efficiently through the strong force. Most importantly, this includes interactions between cosmic-ray protons and nuclei with gas, which can result in the production of one or more pions. The cross sections for pion production can be quite large, comparable to σ_t, but only if the cosmic ray is energetic enough to exceed the threshold for these processes. In particular, in order to produce one

or more pions, a cosmic-ray proton colliding with a proton at rest must have a total energy in excess of $E_p > m_p + 2m_\pi + m_\pi^2/2m_p \approx 1.2$ GeV (see Sec. 3.6).

Pions are unstable, and rapidly decay as follows:

$$\begin{aligned} \pi^0 &\rightarrow \gamma\gamma, \\ \pi^+ &\rightarrow \mu^+ + \nu_\mu \rightarrow e^+ \bar{\nu}_\mu \nu_e \nu_\mu, \\ \pi^- &\rightarrow \mu^- + \bar{\nu}_\mu \rightarrow e^- \nu_\mu \bar{\nu}_e \bar{\nu}_\mu. \end{aligned} \tag{5.23}$$

The decays of neutral pions produce a spectrum of gamma rays that can be directly related to the spectrum of the pions themselves:

$$\frac{dN_\gamma}{dE_\gamma}(E_\gamma) = 2 \int_{E_\pi^{\rm min}(E_\gamma)}^{\infty} \frac{dN_\pi}{dE_\pi}(E_\pi) \, \frac{dE_\pi}{\sqrt{E_\pi^2 - m_\pi^2}}, \tag{5.24}$$

where $E_\pi^{\rm min} = E_\gamma + (m_\pi^2/4E_\gamma)$. For a single decaying pion of energy E_π, this reduces to

$$\begin{aligned} \frac{dN_\gamma}{dE_\gamma}(E_\gamma) &= 2 \int_{E_\pi^{\rm min}(E_\gamma)}^{\infty} \delta(x - E_\pi) \, \frac{dx}{\sqrt{x^2 - m_\pi^2}}, \\ &= \begin{cases} \dfrac{2}{\sqrt{E_\pi^2 - m_\pi^2}}, & \frac{1}{2}\,(E_\pi - \sqrt{E_\pi^2 - m_\pi^2}) < E_\gamma < \frac{1}{2}\,(E_\pi + \sqrt{E_\pi^2 - m_\pi^2}), \\ 0, & \text{otherwise.} \end{cases} \end{aligned} \tag{5.25}$$

In the rest frame of a neutral pion, the decay results in two photons of energy, $E_\gamma = m_\pi/2$. In the lab frame, however, such decays generate a spectrum of gamma rays that is flat between $(E_\pi - \sqrt{E_\pi^2 - m_\pi^2})/2$ and $(E_\pi + \sqrt{E_\pi^2 - m_\pi^2})/2$, and that is zero outside of this range of energies. One implication of this result is that *any* spectrum of neutral pions will produce a gamma-ray spectrum (presented in dN_γ/dE_γ units) that peaks at an energy of $m_\pi/2$.

To make practical use of Eq. 5.24, we will need an expression for the spectrum of pions. Such pions can be produced through the collisions of a hadronic cosmic ray with gas, or with a photon, the latter of which are known as photo-hadronic processes. For collisions with gas, the pion spectrum can be written in terms of the spectrum of cosmic-ray protons (or nuclei), the number density of targets, and the differential cross section for pion production:

$$\frac{dN_\pi}{dE_\pi}(E_\pi) = 4\pi \; n_{\rm gas} \; \eta \int_0^{\infty} dE_p \, \frac{dN_p}{dE_p}(E_p) \, \frac{d\sigma_\pi}{dE_\pi}(E_p, E_\pi), \tag{5.26}$$

where $n_{\rm gas}$ is the number density of target nucleons, and $\eta \sim 1.6$ is a factor that accounts for scattering with heavier nuclei. The quantity $d\sigma_\pi/dE_\pi$ is the differential cross section for proton-proton collisions multiplied by the multiplicity of pions that are produced in the interaction.

The total cross section for inelastic (i.e., pion producing) proton-proton collisions has only a mild dependance on energy, gradually increasing from $\sigma_{pp}^{\rm inel} \approx (3-4) \times 10^{-26}$ cm^2 at $E_p \lesssim 10$ GeV, to $\sigma_{pp}^{\rm inel} \approx 10^{-25}$ cm^2 near $E_p \sim 1$ EeV (see Figure 5.2). For photo-hadronic interactions, in contrast, hadronic resonances make the behavior of the pion production cross section more complex (see Figure 5.3). In the center-of-momentum frame, the total

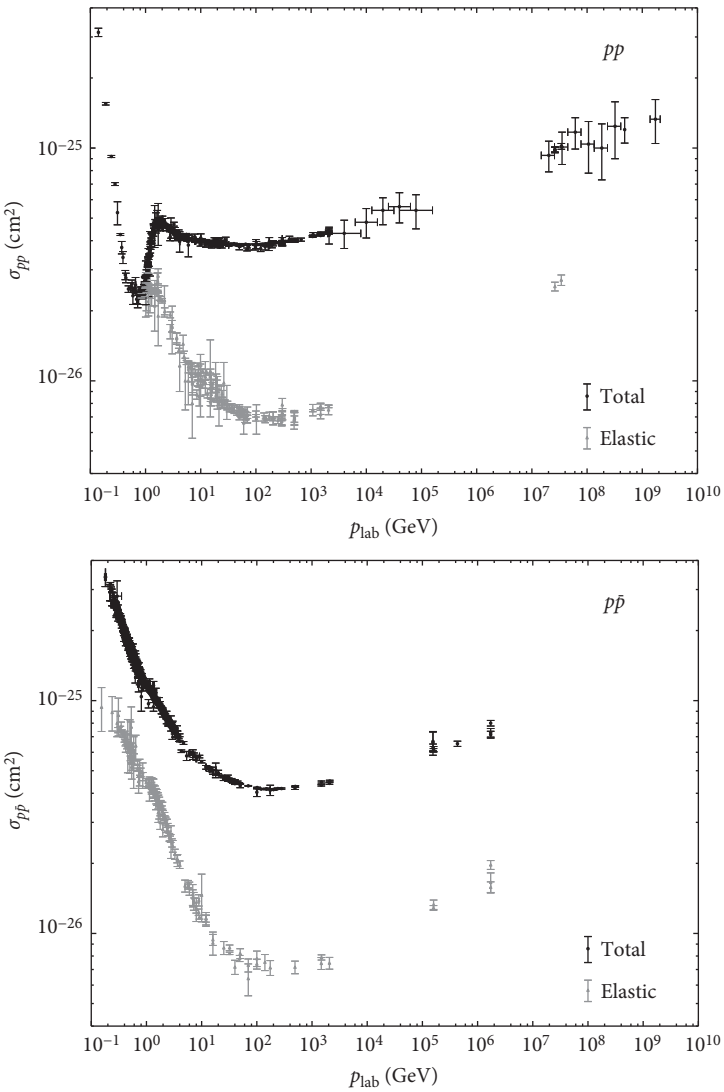


Figure 5.2 Measurements of the total proton-proton (top) and proton-antiproton (bottom) cross sections, along with the corresponding cross sections for elastic scattering [24]. The cross section for inelastic (i.e., pion-producing) scattering is equal to the difference between these quantities.

energy of a proton colliding with a photon is given by

$$E_{\mathrm{CM}} = \left(m_p^2 + 2E_p\epsilon_\gamma[1-\beta\cos\theta]\right)^{1/2} \tag{5.27}$$

$$\approx 45\,\mathrm{GeV} \times \left(\frac{E_p}{\mathrm{PeV}}\right)^{1/2}\left(\frac{\epsilon_\gamma}{\mathrm{MeV}}\right)^{1/2}\left(\frac{1-\cos\theta}{1}\right)^{1/2} \qquad (E_{\mathrm{CM}} \gg m_p),$$

where E_p and ϵ_γ are the energies of the proton and the photon in the lab frame, $\beta = \sqrt{1 - m_p^2/E_p^2}$ is the velocity of the proton, and θ is the angle between the trajectories

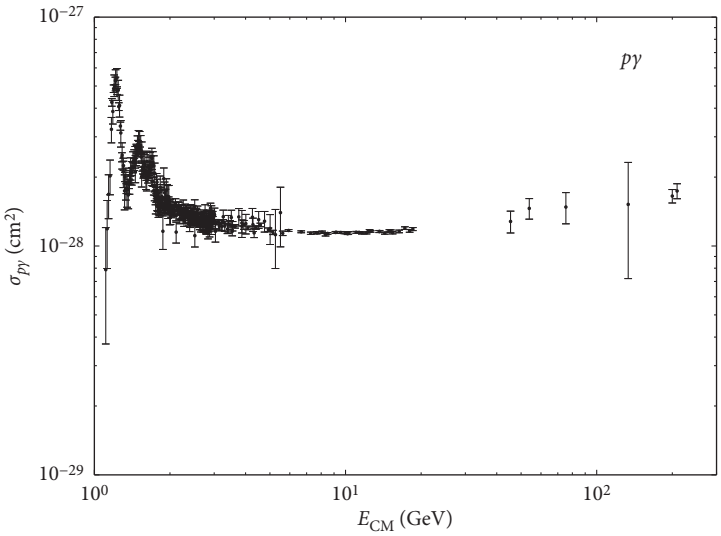


Figure 5.3 Measurements of the total proton-photon (or antiproton-photon) cross section, as a function of the total energy in the center-of-momentum frame [24].

of the two particles. In the last line, I have made the approximation that $E_{\mathrm{CM}} \gg m_p$, which will not always be the case.

For values of E_{CM} in the range of $\sim 1-2\,\mathrm{GeV}$, the production of pions can proceed through a number of distinct hadronic resonances. This can be seen in the energy dependance of $\sigma_{p\gamma}$, as shown in Figure 5.3. The most important of these resonances corresponds to the exchange of a positively charged delta baryon, which has a mass of $m_\Delta \approx 1.232\,\mathrm{GeV}$. For collisions that take place with $E_{\mathrm{CM}} \sim m_\Delta$, this resonance leads to the efficient production of pions through the following processes:

$$\begin{aligned} p + \gamma &\rightarrow \Delta^+ \rightarrow p + \pi^0, \\ p + \gamma &\rightarrow \Delta^+ \rightarrow n + \pi^+, \end{aligned} \tag{5.28}$$

where approximately 2/3 of such interactions produce a neutral pion, while the remaining 1/3 result in the production of a positively charged pion.

While leptonic processes (i.e., bremsstrahlung and inverse Compton scattering) can generate gamma rays, the production of pions in hadronic collisions is unique in that it also leads to the production of high-energy neutrinos. In any conceivable situation in which neutral pions are generated, charged pions (and therefore neutrinos) will be created as well.

There is no simple analytic expression for the differential cross section for pion production. Instead, calculations involving this quantity often make use of one of the code packages in which this information is numerically tabulated, such as Galprop [25], SYBILL [26], QGSJet [27], or EPOS [28].

5.3 Interactions of high-energy photons

Unlike electrons, nucleons, and nuclei, photons are unaffected by magnetic fields and thus travel in straight lines as they propagate through our universe. Furthermore, up to roughly

TeV-scale energies, our universe is largely transparent to photons, making it possible for such particles to reach us from cosmological distances without experiencing any appreciable energy losses (other than those resulting from cosmological redshift). At higher energies, however, gamma rays can be significantly attenuated by scattering with background photons through the process of electron-positron pair production.

Consider a pair of colliding photons with energies of E_γ and ϵ_γ. In the center-of-momentum frame, their total energy is given by $E_{\rm CM} = [2E_\gamma \epsilon_\gamma\ (1 - \cos\theta)]^{1/2}$, where θ is the angle between the incoming photons. In order for these photons to produce an electron-positron pair, their combined energy must exceed twice the electron mass, $E_{\rm CM} > 2m_e$, corresponding to:

$$E_\gamma > \frac{2m_e^2}{\epsilon_\gamma\ (1 - \cos\theta)} \approx 2.6 \times 10^5\ {\rm GeV} \times \left(\frac{10^{-3}\ {\rm eV}}{\epsilon_\gamma}\right)\left(\frac{2}{1 - \cos\theta}\right). \tag{5.29}$$

Thus PeV-scale photons can produce electron-positron pairs by scattering with the CMB. Less energetic gamma rays can also undergo pair production by scattering with photons in the high-energy tail of the CMB, or with other photons, such as those which make up the cosmic infrared background.

The cross section for electron-positron pair production is given by [29, 30, 31]:

$$\sigma_{\gamma\gamma}(E_\gamma, \epsilon_\gamma, \theta) = \frac{3\sigma_t\ (1 - \beta^2)}{16}\left[2\beta(\beta^2 - 2) + (3 - \beta^4)\ln\left(\frac{1 + \beta}{1 - \beta}\right)\right], \tag{5.30}$$

where

$$\beta \equiv \left[1 - \left(\frac{2m_e}{E_{\rm CM}}\right)^2\right]^{1/2}. \tag{5.31}$$

From this cross section, we can integrate over a distribution of target photons to determine the optical depth for a gamma ray propagating a distance l through an isotropic radiation field:

$$\tau(E_\gamma, l) = l \int_{-1}^{1} \frac{1 - \cos\theta}{2}\, d(\cos\theta) \int \sigma_{\gamma\gamma}(E_\gamma, \epsilon_\gamma, \theta)\ \frac{dn}{d\epsilon_\gamma}(\epsilon_\gamma)\, d\epsilon_\gamma, \tag{5.32}$$

where $dn/d\epsilon_\gamma$ is the differential number density of target photons. The probability of a photon traveling a distance l without scattering is given by $e^{-\tau}$.

As as result of these interactions, the universe is largely opaque to very high-energy photons, significantly limiting the distance from which we can observe such particles. The electron-positron pairs that are produced through these interactions, however, go on to experience inverse Compton scattering which can generate a potentially observable flux of lower energy gamma rays (see Chapter 16).

Additional reading

When I teach graduate-level courses on the subject of high-energy astrophysics, I have often used Longair's well-known textbook on this subject. I also recommend the review of cosmic rays included in the Particle Data Group's *Review of Particle Physics* [24].

Problems

5.1 Show that Eq. 5.5 is equivalent to the more familiar form of the expression for bremsstrahlung energy losses:

$$\frac{dE_e}{dt} \approx -\frac{Z(Z+1.3)e^6 n E_e}{16\pi^3 \epsilon_0^3 m_e^2 c^4 \hbar}\left[\ln\left(\frac{183}{Z^{1/3}}\right)+\frac{1}{8}\right], \tag{5.33}$$

where n is the number density of targets of charge Z and ϵ_0 is the permittivity of free space.

5.2 Use energy and momentum conservation to show that for a head-on collision between an electron with energy, E_e, and a photon of energy, ϵ_γ, the maximum energy that the photon can leave the collision with is given by $E_\gamma^{\rm max} = 4E_e^2\epsilon_\gamma/m_e^2$.

5.3 Show that a relativistic electron in the Thomson regime loses half of its energy over a period of time that is related as follows to the energy loss rate: $\tau_{1/2} = E_e\,(dE_e/dt)^{-1}$.

5.4 What is the timescale for energy losses via inverse Compton scattering for a 1 TeV electron in the current universe, at $z = 10$, and at $z = 100$? Don't forget to account for Klein-Nishina suppression.

5.5 Consider an environment in which electrons can scatter with both the CMB ($\rho_{\rm CMB} \approx 0.26\,{\rm eV/cm^3}$, $T_{\rm CMB} \approx 0.00023\,{\rm eV}$) and with starlight ($\rho_{\rm star} \approx 0.6\,{\rm eV/cm^3}$, $T_{\rm star} \approx 0.4\,{\rm eV}$). At what value of E_e is the energy loss rate from scattering with the CMB approximately equal to that from scattering with starlight?

5.6 For a monotonic spectrum of target radiation, find the value of E_γ at which the spectrum of inverse Compton emission, dN_γ/dE_γ, is maximized (as a function of E_e and ϵ_γ). Use this to show that, in the Klein-Nishia regime, this spectrum peaks at an energy that is only slightly below that of the parent electron.

5.7 Exploiting the Lorentz invariance of the square of the energy-momentum four-vector, derive Eq. 5.27.

5.8 Consider a photon that strikes a proton at rest. What minimum energy must this photon have in order to produce a positively charged delta baryon, $\gamma + p \rightarrow \gamma + \Delta$? Recall that $m_\Delta \approx 1.232\,{\rm GeV}$.

5.9 Considering that cosmic rays strike earth's atmosphere with energies as high as $\sim 10^{11}\,{\rm GeV}$, estimate the heaviest possible mass that a particle could have and still be created in such a collision.

5.10 Consider a 1 TeV proton in an environment with a gas density of $n = 1\,{\rm cm^{-3}}$. Estimate how long, on average, it would take for this particle to produce one or more pions. Compare this to the energy loss timescales for the processes of synchrotron and inverse Compton scattering. Remember that synchrotron and inverse Compton scattering are each less efficient for protons than for electrons.

II

Cosmology

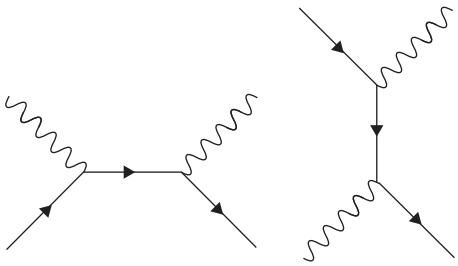
6

Recombination and Photon Decoupling

During the first few hundred thousand years after the Big Bang, the temperature of our universe was high enough to prevent electrons from becoming bound to protons or other nuclei. In this state, all of space will filled with a dense and highly opaque plasma of electrically charged particles. As the temperature dropped, however, electrically neutral atoms began to form, causing our universe to become largely transparent. It was during this transition, known as recombination, that the cosmic microwave background (CMB) was formed.[1] In this short chapter, I describe how the abundance of free electrons evolved over time, and use this information to determine when our universe's photons decoupled from the electrons and other charged particles to become the light that is the CMB.

6.1 Photon scattering

Photons can elastically scatter off of free electrons and other electrically charged particles through the following diagrams:



Following the procedure described in Chapter 3, we can use these diagrams to make a rough estimate for the scattering cross section associated with this process. For two vertex

[1] It has never made much sense to me that we call this transition "recombination." After all, it's not as if the electrons had been previously bound to nuclei during some earlier period of cosmic history. That being said, recombination is now the standard terminology and we seem to be stuck with it.

factors of $\sim g_2 \sin\theta_W = (4\pi\alpha)^{1/2}$ and a propagator factor of $\sim 1/m_e^2$, we obtain an estimate of $\mathcal{M} \sim 8\pi\alpha E_{\rm tot}^2/m_e^2 \sim 8\pi\alpha$ for the amplitude. Combining this with Eq. 3.15, we estimate that this cross section should be of the order of $\sigma \sim 4\pi\alpha^2/m_e^2$. The full tree-level calculation yields the more precise result of

$$\begin{aligned}\sigma_t &= \frac{8\pi\alpha^2}{3m_e^2} \\ &\approx 6.652 \times 10^{-25}\ \text{cm}^2,\end{aligned} \tag{6.1}$$

which is known as the Thomson cross section. Notice that since this cross section scales with m_e^{-2}, we can safely neglect the scattering of photons with protons, which is further suppressed by a factor of $(m_e/m_p)^2 \sim 3 \times 10^{-7}$ relative to that with electrons.

The scattering rate of a photon is given by $\Gamma = n_e \sigma_t c$, where n_e is the number density of free electrons and c is the speed of light. If we naïvely assume that none of the electrons were bound to nuclei during this period, we can relate n_e to the measured baryon density as follows:

$$\begin{aligned}n_e &= n_b \left(1 - \frac{Y_p}{2}\right) \\ &= \frac{n_{b,0}}{a^3}\left(1 - \frac{Y_p}{2}\right),\end{aligned} \tag{6.2}$$

where $Y_p \equiv 4n_{\rm He}/(n_{\rm H} + 4n_{\rm He}) \approx 0.245$ is the fraction of the universe's baryonic mass that is in the form of helium (i.e., the helium mass fraction), and n_b is the number density of baryons, which is related to the current value by $n_b = n_{b,0}\, a^{-3}$, where $n_{b,0} \approx 2.5 \times 10^{-7}\ \text{cm}^{-3}$.

To determine whether the photons will remain bound to the plasma of charged particles, we compare the scattering rate to the rate of Hubble expansion:

$$\begin{aligned}\frac{\Gamma}{H} &\approx \frac{n_{b,0}\, a^{-3}(1 - Y_p/2)\, \sigma_t c}{H_0 \Omega_M^{1/2} a^{-3/2}}, \\ &\approx \frac{3.4 \times 10^{-3}}{a^{3/2}},\end{aligned} \tag{6.3}$$

where I have used $H \approx H_0 \Omega_M^{1/2} a^{-3/2}$, as is appropriate during our universe's matter-dominated era. Under these assumptions, we find that the scattering rate would have exceeded the rate of Hubble expansion for all $a \lesssim 0.023$ (corresponding to redshifts of $z \gtrsim 43$), causing the photons to remain tightly coupled to the electrons up to that point in time. This was not, however, how this process actually played out. In fact, the photons in our universe decoupled from the electrons and other matter around $z \sim 1090$. This occurred so early because most of the electrons had become bound to nuclei during this era, and thus did not significantly contribute to the scattering rate of photons.

To determine the point in time at which photon decoupling took place, we must take into account the ionization history of our universe. In the following section, we'll use the Boltzmann equation to calculate how the fraction of free electrons evolved over time, and use this information to determine when the photons decoupled from our universe's charged particles.

6.2 The ionization history of the early universe

During the period of recombination, the electrons, protons, atoms, and photons that were present in our universe were all maintained in kinetic equilibrium, with a common temperature. Furthermore, since the binding energy of an electron to a helium nucleus (24.6 eV and 54.4 eV) is significantly larger than that for hydrogen (13.6 eV), neutral helium atoms formed first, after which they played little role in the scattering of photons. With this in mind, we'll focus in this section on the fraction of hydrogen that was ionized as a function of temperature, assuming that the helium was already fully neutral.

Throughout this process, the electrons, protons, and hydrogen atoms were all highly nonrelativistic. This allows us to use Eq. 4.8 to relate the number density of neutral hydrogen atoms to the number densities of free protons and free electrons:

$$\begin{aligned} n_{\rm H} &= g_{\rm H}\left(\frac{m_{\rm H}T}{2\pi}\right)^{3/2} e^{-(m_{\rm H}-\mu_{\rm H})/T} \\ &= g_{\rm H}\left(\frac{m_{\rm H}T}{2\pi}\right)^{3/2} e^{-(m_{\rm H}-\mu_p-\mu_e)/T} \\ &= g_{\rm H}\left(\frac{m_{\rm H}T}{2\pi}\right)^{3/2} e^{(m_p+m_e-m_{\rm H})/T}\, e^{-(m_p-\mu_p)/T}\, e^{-(m_e-\mu_e)/T} \\ &= g_{\rm H}\left(\frac{m_{\rm H}T}{2\pi}\right)^{3/2} e^{(m_p+m_e-m_{\rm H})/T}\, \frac{n_p}{g_p}\left(\frac{2\pi}{m_pT}\right)^{3/2}\, \frac{n_e}{g_e}\left(\frac{2\pi}{m_eT}\right)^{3/2} \\ &= n_p n_e \left(\frac{2\pi\, m_{\rm H}}{m_p m_e T}\right)^{3/2} e^{B/T}, \end{aligned} \tag{6.4}$$

where $g_{\rm H} = 4$, $g_p = g_e = 2$, and $B = m_p + m_e - m_{\rm H} = 13.6$ eV is the binding energy of hydrogen. In the second line, I have used the fact that the process $p + e \leftrightarrow {\rm H} + \gamma$ is efficient to relate the chemical potentials according to $\mu_p + \mu_e = \mu_{\rm H}$.

We would like to use this expression to calculate the fraction of hydrogen that is ionized, $X_e \equiv n_p/(n_p + n_{\rm H})$, as a function of temperature. To this end, we can use the fact that the universe is electrically neutral, $n_p = n_e$, and relate the number densities of free protons and hydrogen atoms to the number density of baryons, $n_{\rm H} + n_p = n_b(1 - Y_p)$, to rewrite Eq. 6.4 as follows:

$$\begin{aligned} \frac{1 - X_e}{X_e^2} &= n_b\,(1 - Y_p)\left(\frac{2\pi\, m_{\rm H}}{m_p m_e T}\right)^{3/2} e^{B/T} \\ &= n_{b,0}\,(1 - Y_p)\left(\frac{2\pi\, m_{\rm H} T}{m_p m_e T_{\rm CMB}^2}\right)^{3/2} e^{B/T}, \end{aligned} \tag{6.5}$$

where in the second line I have related the baryon density to the temperature and the present-day temperature of the CMB, $n_b = n_{b,0}\, a^{-3} = n_{b,0}\,(T/T_{\rm CMB})^3$, where $T_{\rm CMB} = 2.35 \times 10^{-4}$ eV. The expression given in Eq. 6.5 is known as the Saha equation for the ionization fraction of hydrogen. Once the value of $(1 - X_e)/X_e^2$ is determined for a given value of the temperature, X_e can be calculated by a straightforward application of the quadratic formula.

The Saha equation can be used to determine the equilibrium value of X_e. As the temperature drops and the number density of free electrons becomes small, however, the rate of recombination becomes comparable to, and eventually lower than, the rate of Hubble expansion, causing the value of X_e to freeze out and depart from its equilibrium value. In this case, we must solve the appropriate Boltzmann equation for $n_e = X_e n_{b,0} a^{-3}(1 - Y_p)$. Following from Eq. 4.34, the evolution of n_e is described by:

$$\begin{aligned}\frac{dn_e}{dt} &= -3Hn_e + n_e^{\rm Eq}\, n_p^{\rm Eq} \langle \sigma\, v \rangle \left[\frac{n_{\rm H}\, n_\gamma}{n_{\rm H}^{\rm Eq}\, n_\gamma^{\rm Eq}} - \frac{n_e\, n_p}{n_e^{\rm Eq}\, n_p^{\rm Eq}} \right] \\ &\approx -3Hn_e + \langle \sigma\, v \rangle \left[\left(\frac{m_e T}{2\pi} \right)^{3/2} [n_b(1 - Y_p) - n_e]\, e^{-B/T} - n_e^2 \right] \\ &\approx -3Hn_e + \langle \sigma\, v \rangle \left[(n_e^{\rm Eq})^2 - n_e^2 \right].\end{aligned} \tag{6.6}$$

The quantity $\langle \sigma\, v \rangle$ in this expression is the thermally averaged cross section for the process $p + e \leftrightarrow {\rm H} + \gamma$. In the second line, we have made use of the facts that $n_\gamma \approx n_\gamma^{\rm Eq}$ and $m_p \approx m_{\rm H}$. In the final line, I have rewritten the result in terms of the equilibrium value of n_e, as determined using the Saha equation in the $X_e \ll 1$ limit.

This calculation is complicated by the fact that any recombination event that produces a hydrogen atom in the ground state will release a photon with enough energy to almost immediately ionize another atom, resulting in no net decrease in n_e or X_e. Furthermore, when an electron that is in a state with a principle quantum number of 3 or more transitions into the ground state of hydrogen, a photon is released with enough energy to excite another hydrogen atom to the $n = 2$ excited state, again preventing any net increase in the number of atoms in the ground state. In light of these considerations, the ground state of hydrogen is most efficiently obtained through the process of an electron and a proton combining into an excited state, followed by a cascade into the $n = 2$ state, and the subsequent transition from the $2s$ to $1s$ state. Importantly, this last step releases not one, but two photons, each with too little energy to either excite or ionize a hydrogen atom (the $2p$ to $1s$ transition, in contrast, releases only a single photon). With this in mind, we will use an effective cross section to describe the recombination rate during this era, which can be parameterized as follows:

$$\langle \sigma\, v \rangle \approx 3.9 \times 10^{-13}\, {\rm cm}^3/{\rm s} \times \left(\frac{\rm eV}{T} \right)^{0.6}. \tag{6.7}$$

For a more detailed treatment, see Ref. [32].

The equilibrium value of the ionization fraction is shown in Figure 6.1, as well as the evolution that results from solving the Boltzmann equation described above. The value of X_e froze out of equilibrium when the Hubble rate exceeded the rate of recombination, $H \gtrsim n_e^{\rm Eq} \langle \sigma\, v \rangle$, which occurred at a temperature of about $T \sim 0.3$ eV.

6.3 Reionization

Although nearly all of the charged particles in our universe became bound together into neutral atoms by $z \sim 1{,}000$, it has not since remained in that configuration. By the time that our universe was two hundred million years old or so ($z \sim 18$), objects had formed that were capable of producing photons with enough energy to ionize hydrogen and helium atoms.

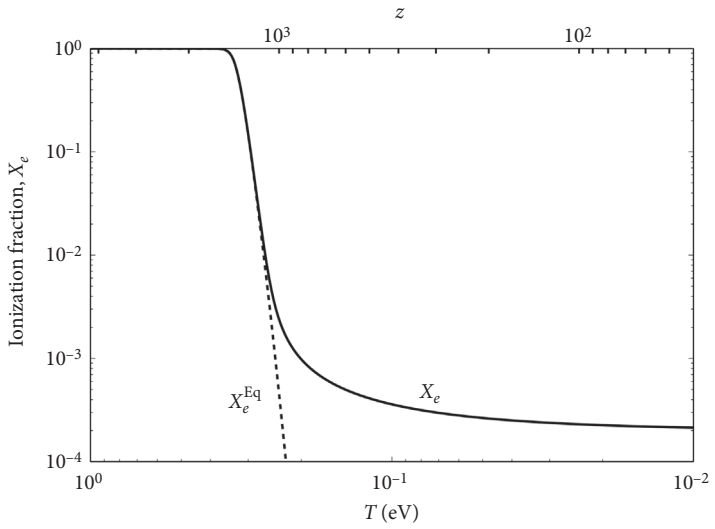


Figure 6.1 The fraction of hydrogen atoms that were ionized as a function of temperature. At early times ($T \gtrsim 0.3$ eV), the ionization fraction closely traced the equilibrium value, as obtained using the Saha equation. As the recombination rate fell below the rate of Hubble expansion, however, the value of X_e froze out of equilibrium and approached an approximately constant value.

While it remains debated precisely which classes of objects were primarily responsible for producing this ionizing radiation, some of the likely candidates include dwarf galaxies, quasars, and population III stars. The radiation from these sources gradually reionized the hydrogen in our universe, causing it to transition from a largely neutral state at $z \sim 20$, to a nearly fully ionized state by $z \sim 6$.

Our most direct information regarding the reionization of our universe comes from measurements of the spectra of high-redshift quasars. Being among the brightest objects in our universe, quasars can be observed from very great distances. During its propagation, the light from these objects can be efficiently absorbed by neutral hydrogen in the intergalactic medium, in particular through Lyman transitions (i.e., those from $n = 1$ to $n \geq 2$ states). The Lyman-alpha transition (from $n = 1$ to $n = 2$) leads to an absorption line at an observed wavelength of $\lambda \approx (121.6\,\text{nm})/(1+z)$, where z is the redshift of the encountered hydrogen. Since the light from a distant quasar passes through clouds of neutral hydrogen at a variety of redshifts over its line of sight, the observed spectrum from such a source will be suppressed over a range of wavelengths, leading to the spectral feature known as the Gunn-Peterson trough. From observations of this feature in the light from high-redshift quasars, it is clear that the hydrogen found in the intergalactic medium is almost entirely ionized between $z \sim 6$ and the present.

At higher redshifts, which cannot be readily probed using quasar absorption lines, our knowledge of the epoch of reionization is informed in large part by measurements of the total integrated optical depth of the CMB, τ. The scattering of CMB photons with free electrons can damp the CMB anisotropies, leading to an approximately scale-independent suppression of the observed angular power spectrum above $l \approx 10$ (see Chapter 7) by a factor of $e^{2\tau}$, where

$$\tau = \int n_e \sigma_t c \, dt, \tag{6.8}$$

and where n_e is the number density of free electrons. In other words, this scattering reduces the already small variations in the temperature of the CMB, in particular in those patches of the sky that are $\sim 10^\circ - 20^\circ$ or smaller in extent.

We can use $H \equiv \dot{a}/a$ to rewrite Eq. 6.8 in terms of the scale factor,

$$\tau(a_{\rm min}) = \int_{a_{\rm min}}^{1} \frac{n_e(a)\sigma_t c}{aH(a)}\, da, \tag{6.9}$$
$$= \int_{a_{\rm min}}^{1} \frac{n_{b,0}(1-Y_p)X_e(a)\sigma_t c}{a^4 H(a)}\, da,$$

or in terms of redshift,

$$\tau(z_{\rm max}) = -\int_{z_{\rm max}}^{0} \frac{n_e(z)\sigma_t c}{(1+z)H(z)}\, dz, \tag{6.10}$$
$$= \int_{0}^{z_{\rm max}} \frac{n_{b,0}(1-Y_p)X_e(z)\sigma_t c\,(1+z)^2}{H(z)}\, dz.$$

The final data release by the Planck Collaboration reports a value of $\tau \approx 0.054$ integrated below $z_{\rm max} = 50$. If we adopt a simple step function model, with $X_e = 1$ below a redshift $z_{\rm re}$ and $X_e = 0$ otherwise, this measurement implies that our universe became reionized at $z_{\rm re} \sim 8.5$. More realistically, our universe was likely reionized gradually over a period of time between $z \sim 20$ and $z \sim 6$.

6.4 The surface of last scattering

Now that we have worked out the ionization history of our early universe, we can return to the question of when the photons decoupled from the electrons and other charged particles to form the CMB.

Consider a photon that has been traveling through our universe since the era of recombination. A photon will, on average, have scattered the following number of times since being emitted at t_0:

$$\tau(t_0) = \int_{t_0}^{t_{\rm now}} n_e \sigma_t c\, dt. \tag{6.11}$$

In this context, τ is referred to as the optical depth of the universe. As we did in the previous section, we can use $H \equiv \dot{a}/a$ to rewrite this expression in terms of redshift:

$$\tau(z_0) = -\int_{z_0}^{0} \frac{n_e(z)\sigma_t c}{(1+z)H(z)}\, dz, \tag{6.12}$$
$$= \int_{0}^{z_0} \frac{n_{b,0}(1-Y_p)X_e(z)\sigma_t c\,(1+z)^2}{H(z)}\, dz.$$

Using the ionization history of our universe, $X_e(z)$, as described in the past two sections, and for an expansion rate given by $H(z) = H_0[\Omega_M(1+z)^3 + \Omega_\Lambda + \Omega_{\rm rad}(1+z)^4]^{1/2}$, we find that $\tau = 1$ for $z_0 \approx 1{,}090$, corresponding to a time 380,000 years after the Big Bang.

When we detect and measure the CMB, we are observing photons that have been traveling without significant scattering or deflection since $z \sim 1090$. The CMB, as we observe it, is an image of the surface of a sphere that surrounds us, with a radius corresponding to $z \approx 1{,}090$. This is known as the surface of last scattering, and it represents the point at which the typical CMB photon was most recently scattered (or equivalently, the time at which our universe became transparent to photons).

Additional reading

For a more detailed discussion of recombination and photon decoupling, I recommend Weinberg's *Cosmology*.

Problems

6.1 Estimate the Thomson scattering rate for a photon propagating through the Milky Way. What fraction of photons emitted from the Galactic Center region scatter en route to the solar system?

6.2 In Sec. 6.2, I asserted that our universe is electrically neutral, meaning that it contains effectively zero net electric charge. Calculate the electrostatic force that would have been felt by an electron at a distance of 1 AU from the sun if the sun contained 0.0001% more protons than electrons. How small would such a charge asymmetry have to be in order for the magnitude of the gravitational force felt by this electron to be equal to that of the corresponding electrostatic force?

6.3 To obtain a precise solution to the Saha equation (Eq. 6.5), one generally applies the quadratic formula. Instead, find an expression for $X_e^{\rm Eq}(T)$ in the $X_e \ll 1$ limit. How does $X_e^{\rm Eq}$ scale with the temperature in this regime?

6.4 Estimate the residual value of X_e long after decoupling by determining the redshift and the value of X_e when the recombination rate was equal to the rate of Hubble expansion. How does this post-freeze-out value of X_e depend on the cross section for recombination?

6.5 Write a code to numerically solve the Boltzmann equation given in Eq. 6.6.

6.6 In Eq. 6.2, I wrote the total number density of electrons as $n_e = n_b\,(1 - Y_p/2)$, while later in this chapter I wrote the number density of free electrons as $n_e = n_b\,(1 - Y_p)X_e$. Why are these expressions not identical in the $X_e = 1$ limit?

6.7 What is the optical depth of the universe out to $z = 6$? Do the calculation in three different ways: assuming that the universe is matter-dominated with $\Omega_M = 1$, assuming that the universe is dark energy-dominated with $\Omega_\Lambda = 1$, and using the measured abundances of matter and dark energy.

6.8 What would the redshift of the surface of last scattering have been if our universe contained only helium and no hydrogen? How would this have impacted the temperature of the CMB?

7

The Cosmic Microwave Background

The CMB has played a central role in the science of cosmology since its accidental discovery in 1965 by Arno Penzias and Robert Wilson. The measured temperature and spectral shape of this radiation helped to establish the Big Bang theory as the leading cosmological paradigm. In recent decades, the detailed temperature variations in the CMB have been a treasure trove to cosmologists, providing us with a vast body of information from which we can deduce the evolution, composition, and geometry of our universe. In this chapter, I review the observed characteristics of the CMB, including its spectrum, as well as its temperature and polarization angular power spectra. I also discuss how the anisotropies in the CMB came to be and what they can tell us about our universe.

7.1 The spectrum of the CMB

The spectrum of the CMB is a nearly perfect thermal blackbody. In fact, it more closely resembles an idealized blackbody than any other source or object that has ever been observed. This arises from the fact that, prior to their decoupling, these photons were in a state of nearly perfect kinetic equilibrium. As a result, the distribution of these photons' energies at the time of decoupling was very well described by an equilibrium distribution (see Eq. 4.4),

$$\frac{dn_\gamma}{dE_\gamma} = \frac{1}{\pi^2} \frac{E_\gamma^2}{e^{E_\gamma/T} - 1}, \tag{7.1}$$

where T, in this context, is the temperature of the photons, electrons, and atoms at the time that the CMB was being formed, during the era of recombination (or equivalently, near the surface of last scattering). As our universe subsequently expanded, these photons lost energy through the effects of cosmological redshift, and their number density was diluted. As discussed in Sec. 4.6, a decoupled population of relativistic particles remains in an equilibrium distribution as the universe expands, with the temperature of that distribution evolving as $T \propto 1/a$. Thus, as our universe expanded, the spectrum of the CMB continued to be that of a blackbody, with a temperature that continuously decreased over time.

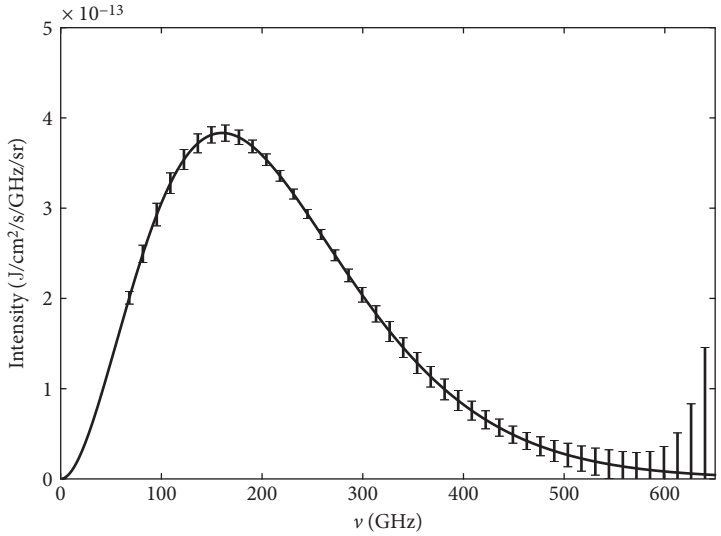


Figure 7.1 The intensity of a $T = 2.7255$ K blackbody (solid curve), compared to the COBE satellite's measurements of the spectrum of the CMB (error bars) [33]. In order to make these error bars visible, I have exaggerated their size by a factor of 500.

Note that the equilibrium distribution of photons given in Eq. 7.1 is equivalent to the familiar Planck blackbody spectrum:

$$B_\nu = \frac{4\pi\,\nu^3}{e^{2\pi\nu/T} - 1}, \tag{7.2}$$

where B_ν is the spectral radiance, defined as the energy per time, per area, per solid angle, per frequency.

Integrated across all directions of the sky, the average temperature of the CMB in our universe today has been measured to be

$$\langle T \rangle = \frac{1}{4\pi} \int T(\theta, \phi)\ \sin\theta\, d\theta\, d\phi \approx 2.7255\ \text{K}. \tag{7.3}$$

At this temperature, the average photon in the CMB has an energy of $\langle E_\gamma \rangle \approx 6.3 \times 10^{-4}$ eV, and a wavelength of 0.3 mm.

In Figure 7.1, the spectrum of the CMB (as measured by the FIRAS instrument onboard the COBE satellite) is compared to that of a 2.7255 K blackbody. In plotting these error bars, I have exaggerated their size by a factor of 500. Without this artificial enhancement, these error bars would have been invisible, entirely obscured by the line width of the solid curve. Measurements such as these allow us to constrain the temperature of this blackbody to remarkable precision, $T = 2.72548 \pm 0.00057$ K [34]. The spectrum of the CMB is indeed a spectacular example of a nearly perfect, and remarkably well measured, blackbody.

By integrating Eq. 7.1, we can calculate the number density and energy density of the photons that make up the CMB:

$$n_\gamma = \frac{1}{\pi^2}\int \frac{E_\gamma^2\, dE_\gamma}{e^{E_\gamma/T}-1} \approx 411\,\mathrm{cm}^{-3}, \tag{7.4}$$

$$\rho_\gamma = \frac{1}{\pi^2}\int \frac{E_\gamma^3\, dE_\gamma}{e^{E_\gamma/T}-1} \approx 0.260\,\mathrm{eV/cm}^3.$$

Comparing this result to the critical density, we find $\Omega_\gamma \equiv \rho_\gamma/\rho_{\rm crit} \approx 5.3\times10^{-5}$. Despite its importance to cosmologists, the CMB does not make up much of our universe's current energy density.

7.2 Temperature anisotropies

The temperature of the CMB is not identically equal to its average value in all directions of the sky. We can define these temperature variations as follows:

$$\frac{\delta T}{T}(\theta,\phi) \equiv \frac{T(\theta,\phi)-\langle T\rangle}{\langle T\rangle}. \tag{7.5}$$

The largest source of these temperature variations comes from the Doppler shift associated with the earth's motion relative to the CMB's rest frame, which induces a dipole anisotropy with a magnitude of $\delta T \approx 3.36$ mK. After subtracting this cosmologically uninteresting dipole anisotropy, the root mean square of the remaining variations in the temperature of this radiation has been measured to be

$$\left\langle \left(\frac{\delta T}{T}\right)^2 \right\rangle^{1/2} \approx 1.1\times10^{-5}, \tag{7.6}$$

corresponding to typical temperature variations on the scale of $\delta T \approx 30\,\mu$K. These temperature anisotropies were first detected by the COBE satellite in 1992, and have since been studied with steadily increasing precision and over a much wider range of angular scales. To the best of our current ability to measure, the temperatures of the various regions of the CMB are distributed according to a Gaussian distribution, centered at $\langle T\rangle$ and with a width given by $\sigma = \langle \delta T^2\rangle^{1/2}$.

A number of different physical mechanisms have contributed to the generation of the CMB's temperature anisotropies. Although we will discuss these in more detail later, I will briefly mention here the most important of these mechanisms, known as the Sachs-Wolfe effect. The Sachs-Wolfe effect arises from the fact that the energy density of our universe at the time of recombination was not perfectly uniform. A photon that originates from a region of higher than average energy density will lose energy through gravitational redshift as it escapes from that region. Similarly, photons from low density regions will gain energy as they move away from those regions. As a result of this effect, the CMB contains hot and cold spots that map onto the regions of low and high density at the time of last scattering.

7.3 The angular power spectrum of the CMB

To compare the predictions of a given cosmological model to the temperature anisotropies observed in the CMB, we first need a way of quantitatively describing the information encoded in this data. This does not mean that we need to know the location, magnitude,

Figure 7.2 The temperature anisotropies of the CMB (after subtracting the dipole), as measured by the Planck satellite. To the best of our ability to measure, these temperatures follow a Gaussian distribution with $\langle T \rangle \approx 2.7255$ K and $\sigma \approx 3.0 \times 10^{-5}$ K. From http://astrog80.astro.cf.ac.uk/HP_bkp/planck.cf.ac.uk

or angular size of every hot or cold spot in the CMB. After all, no cosmological model predicts these kinds of specific features, which depend on numerous accidents of history, including the location that a particular observer happens to occupy in the universe. Instead, the predictions of such models are generally limited to the statistical distribution of the CMB's hot and cold spots, as a function of the angular scale (i.e., the size of the hot and cold spots). If we assume that the temperature anisotropies are distributed in a rotationally symmetric way on the sky, and that they are randomly drawn from a Gaussian distribution, this information is entirely encoded in what is known as the CMB's angular power spectrum.

In this section, we will derive the angular power spectrum of the CMB from the distribution of its temperatures on the sky. To this end, we'll begin by expanding the temperature anisotropies in terms of spherical harmonics, $Y_{lm}(\theta, \phi)$, which constitute a complete basis on the celestial sphere:

$$\frac{\delta T(\theta, \phi)}{T} = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_{lm}\, Y_{lm}(\theta, \phi). \tag{7.7}$$

The spherical harmonics are, in turn, defined in terms of the associated Legendre polynomials, $P_l^m(\cos\theta)$:

$$Y_{lm}(\theta, \phi) \equiv \left(\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!} \right)^{1/2} P_l^m(\cos\theta)\, e^{im\phi}. \tag{7.8}$$

Since each associated Legendre polynomial is a function of only θ, the ϕ dependence in the spherical harmonics comes entirely from the factor of $e^{im\phi}$. Also note that since the values of δT are real, $a_{lm}^\star = a_{l-m}$.

Using the fact that the spherical harmonics constitute a complete set of orthogonal functions that span the unit sphere, we can write the following expression for each a_{lm} (see Problem 7.7):

$$a_{lm} = \int d\hat{n}\, \frac{\delta T(\hat{n})}{T}\, Y_{lm}^\star(\hat{n}), \tag{7.9}$$

where $\hat{n}$ denotes a direction in the sky (i.e., a value of θ and ϕ). Under the very reasonable assumption that the statistical properties of the CMB are invariant under rotation, there should not be any correlation between any two $a_{l,m}$ and $a_{l',m'}$ unless $l = l'$ and $m = -m'$. This allows us to write

$$\langle a_{lm} a_{l'm'} \rangle = \delta_{ll'} \delta_{m-m'} C_l, \tag{7.10}$$

which one can think of as the definition of C_l. If the temperature variations of the CMB are entirely Gaussian, all of the information in this data set that is relevant to cosmology (i.e., that is useful for determining cosmological parameters) will be encoded in the set of all C_ls. In other words, everything we can hope to learn about our universe from the CMB's temperature anisotropies can be expressed in terms of the C_ls.

As it appears in Eq. 7.10, $\langle a_{lm} a_{l'm'} \rangle$ denotes the average value of $a_{lm} a_{l'm'}$ taken over an arbitrarily large ensemble of measurements. One could think of this as an average of the measurements performed by a large number of hypothetical observers located at different positions in our universe. Alternatively, it could be thought to represent an average over many CMBs that we could imagine existing across a large group of universes with the same age, composition, and laws of physics as our own. With this in mind, while we might use a given cosmological model to predict the C_ls, these are clearly not, strictly speaking, observationally accessible quantities. This comes down to the fact that we only have one sky (and one set of a_{lm}) to observe and measure, fundamentally limiting our ability to determine the values of the C_ls. In practice, instead of the C_ls, we observationally determine the closely related quantities, C_l^{obs}, by averaging over different values of m:

$$\begin{aligned} C_l^{\text{obs}} &= \frac{1}{2l+1} \sum_{m=-l}^{l} a_{lm}\, a_{l-m} \\ &= \frac{1}{2l+1} \sum_{m=-l}^{l} |a_{lm}|^2. \end{aligned} \tag{7.11}$$

For large values of l, there are many values of m to average over, and C_l^{obs} approaches the theoretically predicted value of C_l. For small values of l, however, the limited number of regions of that angular size leads to the following mean difference between these quantities:

$$\begin{aligned} \left\langle \left(\frac{C_l^{\text{obs}} - C_l}{C_l} \right)^2 \right\rangle &= \left\langle \left(\frac{1}{(2l+1)C_l} \sum_{m=-l}^{l} a_{lm} a_{l-m} - 1 \right)^2 \right\rangle \\ &= \left\langle \frac{1}{(2l+1)^2 C_l^2} \sum_{m=-l}^{l} a_{lm} a_{l-m} \sum_{m'=-l}^{l} a_{lm'} a_{l-m'} + 1 - \frac{2C_l^{\text{obs}}}{C_l} \right\rangle \\ &= \frac{1}{(2l+1)^2 C_l^2} \sum_{m=-l}^{l} \sum_{m'=-l}^{l} \langle a_{lm} a_{l-m} a_{lm'} a_{l-m'} \rangle - 1 \\ &= \frac{1}{(2l+1)^2 C_l^2} \sum_{m=-l}^{l} \sum_{m'=-l}^{l} \Big[\langle a_{lm} a_{l-m} \rangle \langle a_{lm'} a_{l-m'} \rangle \\ &\quad + \langle a_{lm} a_{lm'} \rangle \langle a_{l-m} a_{l-m'} \rangle + \langle a_{lm} a_{l-m'} \rangle \langle a_{l-m} a_{lm'} \rangle \Big] - 1 \end{aligned} \tag{7.12}$$

$$= \frac{1}{(2l+1)^2 C_l^2}\left[(2l+1)^2 C_l^2 + (2l+1)C_l^2 + (2l+1)C_l^2\right] - 1$$
$$= \frac{2}{2l+1}.$$

This underlying difference between the measured values of the $C_l^{\rm obs}$s and the predicted values of the C_ls is known as cosmic variance, and it provides a fundamental limitation to how much cosmological information we can extract from the temperature anisotropies of the CMB.

For example, even if we were able to *perfectly* measure the value of $C_{l=100}^{\rm obs}$ (corresponding to the temperature variations on an angular scale of $\theta \sim \pi/l \approx 1.8°$), we should still expect this result to differ from the value of $C_{l=100}$ predicted by the correct cosmological model at a level of roughly $\sim (2/201)^{1/2} \sim 10\%$. The reason for this comes down to the fact that the CMB contains a finite number of independent regions of this angular size. Once you have measured all of their temperatures, there is no way to further probe the underlying statistical distribution from which those temperatures were drawn.

Another way to think about the angular power spectrum of the CMB is in terms of the correlation between pairs of points separated by a given angular distance. This is characterized by what is known as the 2-point correlation function:

$$C(\theta) = \frac{1}{T^2}\left\langle \delta T(\hat{n})\, \delta T(\hat{n}') \right\rangle_{\hat{n}\cdot\hat{n}'=\cos\theta}, \tag{7.13}$$

where $\hat{n}$ and $\hat{n}'$ represent directions on the sky, and the large brackets denote the average over all pairs of directions that are separated by an angle θ, such that $\hat{n}\cdot\hat{n}' = \cos\theta$. Combining this expression with Eq. 7.7, we can write

$$C(\theta) = \left\langle \sum_{l=1}^{\infty}\sum_{m=-l}^{l} a_{lm}\, Y_l^m(\hat{n}) \sum_{l'=1}^{\infty}\sum_{m'=-l'}^{l'} a_{l'm'}\, Y_{l'}^{m'}(\hat{n}') \right\rangle_{\hat{n}\cdot\hat{n}'=\cos\theta}. \tag{7.14}$$

Further combining this with Eq. 7.10, we can relate the correlation function to the $C_l's$ as follows:

$$C(\theta) = \left\langle \sum_{l=1}^{\infty}\sum_{m=-l}^{l} a_{lm}\, a_{l-m}\, Y_l^m(\hat{n})\, Y_l^{-m}(\hat{n}') \right\rangle_{\hat{n}\cdot\hat{n}'=\cos\theta} \tag{7.15}$$
$$= \sum_{l=1}^{\infty}\sum_{m=-l}^{l} C_l\, Y_l^m(\hat{n})\, Y_l^{-m}(\hat{n}')$$
$$= \sum_{l=1}^{\infty}\sum_{m=-l}^{l} C_l \frac{(2l+1)}{4\pi} P_l^m(\cos\theta)\, P_l^{-m}(1)$$
$$= \sum_{l=1}^{\infty} C_l \frac{(2l+1)}{4\pi} P_l(\cos\theta),$$

where in the third line I have exploited the rotational invariance of the sky, allowing me to choose an orientation such that $n' = 0$. Then, in the last step, I have used the fact that $\sum_{m=-l}^{l} P_l^m(x) P_l^{-m}(1) = P_l(x)$, where $P_l(x)$ are the Legendre polynomials (see Problem 7.9).

Inverting this expression, we can write each C_l as an integral of the correlation function convolved with the corresponding Legendre polynomial over the sky:

$$C_l = \frac{1}{4\pi} \int d\hat{n}\, d\hat{n}'\, P_l(\cos\theta)\, C(\theta) \tag{7.16}$$

$$= \frac{1}{4\pi\, T^2} \int d\hat{n}\, d\hat{n}'\, P_l(\cos\theta) \left\langle \delta T(\hat{n})\, \delta T(\hat{n}') \right\rangle_{\hat{n}\cdot\hat{n}'=\cos\theta}.$$

In Figure 7.3, I show the angular power spectrum of the CMB, as measured by the Planck satellite, over angular scales of $2 < l < 30$ (top) and $30 < l < 2{,}500$ (bottom), corresponding to 90° to 6° and 6° to 0.072°, respectively. The solid curve in each frame is the prediction made by the best-fit ΛCDM cosmological model, and the grey bands around that curve represent the inherent uncertainty associated with cosmic variance. The detailed bumps and wiggles expressed in this curve contain information about how our universe has evolved and expanded over cosmic time. From the shape of this angular power spectrum, we can determine a great deal about the geometry, composition, and history of our universe.

7.4 The evolution of density perturbations

Gravity, as well as the scattering of particles, can cause the distributions of matter and energy to evolve as time passes and the universe expands. While gravity has the tendency to cause regions of higher than average density to collapse and become more dense, other effects such as diffusion can work against this. A full calculation of the evolution of density perturbations in the early universe can be very involved, and is beyond the scope of this book. Instead, the goal of this and the following sections is to provide the reader with a qualitative description of the most important factors that determine how this evolution plays out, hopefully enabling one to develop some intuition for why the angular power spectrum of the CMB looks the way that it does.

Before moving on, I need to introduce some formalism. In calculating the evolution of density perturbations in the universe, it is common to describe these perturbations in terms of an expansion in Fourier space, writing them as a series of plane waves with different wavelengths. As it turns out, each of these Fourier modes evolves independently of the others, making this formalism quite powerful. As the universe expands, the physical wavelength of each Fourier mode is stretched along with it. The comoving wavelength of each mode, however, remains fixed. For this reason, it is convenient to refer to a given mode in terms of its comoving wavenumber, $k \equiv 2\pi/\lambda$.

Note that the angular power spectrum of the CMB is useful to cosmologists only because each of these individual Fourier modes is in phase. If it were not for this fact, the peaks and troughs that we observe in this spectrum would have been washed out. The coherence of these modes is usually interpreted as evidence that the density perturbations in question were initially larger than the size of the cosmological horizon, and started to evolve only when the expansion rate of the universe became slow enough for such regions to become causally connected [36].

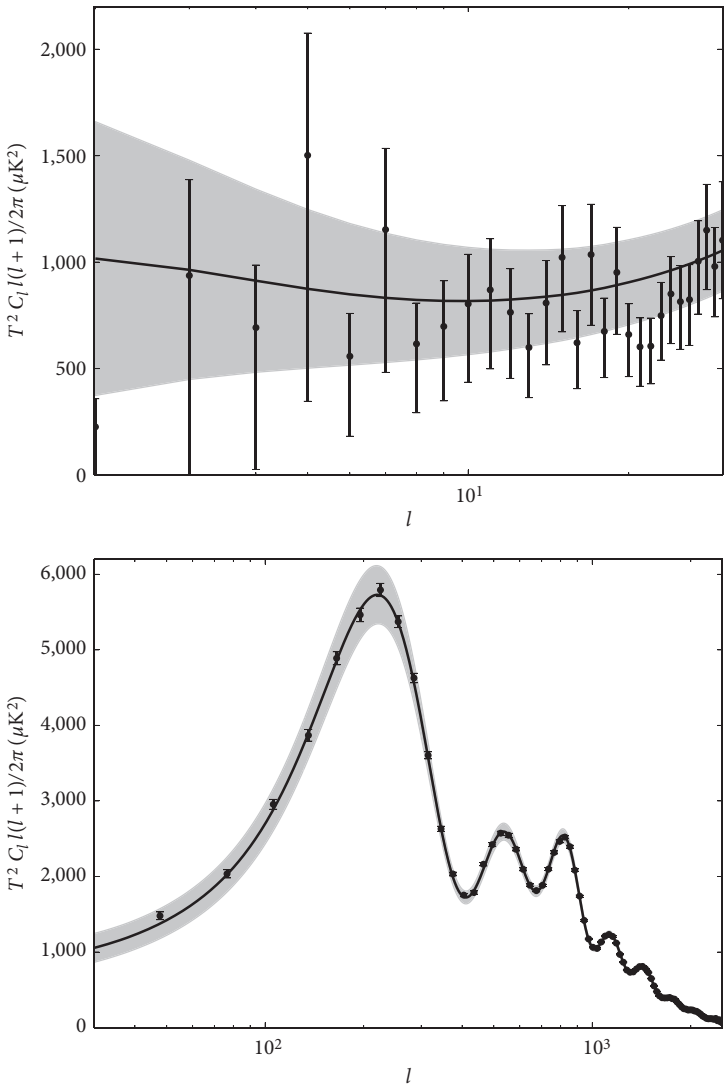


Figure 7.3 The temperature angular power spectrum of the CMB, as measured by the Planck satellite, over angular scales of $2 < l < 30$ (top) and $30 < l < 2{,}500$ (bottom). The solid curve is the prediction made by the best-fit ΛCDM cosmological model, and the grey bands around that curve represent the inherent uncertainty associated with cosmic variance. From the Planck Legacy Archive [35].

Prior to the time of photon decoupling, our universe was filled with energy that behaved as a fluid. In the early universe, regions of higher than average density were compressed by gravity. The photons in the baryon-photon fluid, however, supplied pressure, which pushed back against this compression, leading to oscillations between periods of compression and rarefaction. These waves, which resulted from the interplay between gravity and pressure, were, in fact, sound waves, or acoustic oscillations.

When the photons decoupled from the baryonic matter at the time of recombination, these particles stopped generating pressure, thus bringing an end to the acoustic oscillations.

After a brief period in which scattering caused the photons to move along a random walk, they began to free stream, or travel in straight lines without significant scattering. The Fourier modes whose oscillations were at an extremum at the time of decoupling correspond to the sizes of the regions that were in a state of maximum compression (high density) or rarefaction (low density) at that time. Through the Sachs-Wolfe effect, any photons that were emitted from a region of higher than average density were gravitationally redshifted as they left that region, appearing in the CMB as a patch of colder than average temperature. In contrast, the photons from low density regions at the time of last scattering were blueshifted, and appear as hot spots in the CMB.

7.5 Features of the CMB's angular power spectrum

7.5.1 The Sachs-Wolfe plateau

At the time of recombination, our universe was divided into many different causally disconnected regions, determined by the size of the cosmological horizon. By modifying Eq. 2.58, we can calculate the size of the horizon at the time of last scattering:

$$\begin{aligned} d_h &= a_{\rm ls}\int_0^{t_{\rm ls}}\frac{c\,dt}{a(t)} \\ &= a_{\rm ls}\int_0^{a_{\rm ls}}\frac{c\,da}{H(a)\,a^2} \\ &= \frac{a_{\rm ls}c}{H_0}\int_0^{a_{\rm ls}}\frac{da}{[\Omega_M a+\Omega_{\rm rad}]^{1/2}} \approx 0.24\,{\rm Mpc}. \end{aligned} \tag{7.17}$$

As I will discuss further in Sec. 7.6.1, regions of this size correspond to angular scales of $l\sim 160$, or a little larger than 1°. On scales corresponding to $l \ll 160$, the regions in question were too large to be entirely in causal contact, making it impossible for acoustic waves to have significantly evolved or oscillated on these scales. For this reason, the CMB's angular power spectrum at small values of l directly reflects the initial conditions of the very early universe. These initial conditions are known as the spectrum of primordial density perturbations. Like those on other angular scales, these density perturbations translate into temperature variations in the CMB through the Sachs-Wolfe effect.

It is common to parameterize the spectrum of the primordial density perturbations in terms of a power-law:

$$\mathcal{P}_{\mathcal{R}} = A_s\left(\frac{k}{k_0}\right)^{n_s-1}, \tag{7.18}$$

where A_s is the amplitude of the primordial density perturbations at a particular scale, k_0, and n_s is the spectral index of those perturbations. For the specific choice of $n_s = 1$, the primordial density perturbations are scale-invariant. In other words, they appear the same (statistically speaking), as one zooms in or out of a given region of space. In contrast, a spectrum with $n_s < 1$ will tend to feature greater density perturbations on longer length scales (i.e., smaller k). The spectrum associated with the scale-invariant case of $n_s = 1$ is sometimes called the Harrison-Peebles-Zel'dovich spectrum.

Although we do not know for certain how the initial conditions of our universe were established, most modern cosmologists consider the theory of inflation to be the leading candidate explanation for the generation of the primordial density perturbations (see Chapter 13). Through inflation, quantum fluctuations can generate primordial density perturbations on all scales, with a spectrum that is determined by the energy scale of inflation and the shape of the inflationary potential. A prediction of most theories of inflation is that the spectrum of primordial density perturbations should be approximately scale-invariant, with a spectral index that is only slightly less than unity. Long after these predictions were originally made, they were confirmed by measurements of the CMB's angular power spectrum. In particular, the Planck Collaboration has reported a measurement of $n_s = 0.9649 \pm 0.0042$, corresponding to a nearly scale-invariant primordial spectrum, with only slightly more power on the largest scales.

The impact of the spectrum of primordial density perturbations is particularly evident at $l \lesssim 100$, where the size of the horizon prevented the evolution of any acoustic oscillations. For the case of a perfectly scale-invariant primordial spectrum ($n_s = 1$), one predicts an approximately flat angular power spectrum at low multipoles, $l(l+1)C_l \approx$ constant. This feature, which can be seen in the top frame of Figure 7.3, is known as the Sachs-Wolfe plateau.

7.5.2 The ISW rise

On the largest scales, $l \lesssim 10$, the angular power spectrum of the CMB is enhanced by what is known as the integrated Sachs-Wolfe (ISW) effect. Although these large regions were not in causal contact in the early universe, the gravitational potentials that these photons experienced as they traveled through the universe at much later times did impact this portion of the CMB's angular power spectrum. More specifically, the ISW effect is the result of time variations in the gravitational potentials that these photons traveled through in the low-redshift universe. In cosmological models with a positive equation of state, $w > 0$, the ISW effect suppresses the angular power spectrum of the CMB at the lowest ls. Due to the presence of dark energy, however, our universe's equation of state is negative, causing the lowest C_ls to be subtly enhanced (as seen in the top frame of Figure 7.3). This feature is known as the ISW rise.

7.5.3 The acoustic peaks

The oscillating acoustic waves that resulted from the interplay between gravity and pressure in the early universe led to some of the most distinctive and informative features in the CMB's angular power spectrum. Like the sound waves generated by an organ pipe or guitar string, the spectrum of density perturbations at the time of recombination consisted of an evenly spaced series of frequency harmonics of the fundamental tone. Whereas the length of a musical instrument sets its fundamental tone, the fundamental tone of the CMB's acoustic peaks was determined by the distance that sound waves were able to travel by the time of recombination.

The frequency of the oscillations associated with a given mode is given by $\omega = kc_s$, where k is the wavenumber of the mode, and c_s is the speed of sound. Because the amount of time needed to compress the fluid is proportional to the length scale of the region, oscillations on the largest scales evolve more slowly than those on smaller scales.

When the photons decoupled and began to free stream, the pressure that they had supplied up to that point disappeared, causing the acoustic waves to cease their evolution and freeze into their present state. Each of the acoustic peaks in the CMB's angular power

spectrum appears at a scale that corresponds to a mode of oscillation that was at an extremum at the time of decoupling. This applies both to oscillations that were in a state of maximum or minimum density at the time of recombination; each of these configurations corresponds to enhanced power on a given scale.

Consider the lowest wavenumber mode that was at an extremum at the time of recombination. The wavenumber of this mode is given by $k_1 = \pi/d_s$, where d_s is the sound horizon, or the distance that sound waves can travel by the time of recombination. On the scale corresponding to modes of this wavenumber, regions were reaching a point of maximum compression for the first time just as recombination was taking place. The sound horizon is related as follows to the sound speed, c_s, and the expansion history of the early universe:

$$d_s = a_{\rm ls} \int_0^{a_{\rm ls}} \frac{c_s(a)\, da}{H(a)\, a^2}, \tag{7.19}$$

where $a_{\rm ls}$ is the scale factor at the time of last scattering. Notice that this is similar to the expression for the horizon size at the time of last scattering (see Eq. 7.17), but scaled down by a factor of c_s/c. In a perfect relativistic fluid (without any nonrelativistic baryons), the sound speed is given by $c_s = c/\sqrt{3}$. In more generality, it can be written as $c_s = \sqrt{w}\, c$, where the equation of state, w, depends on the abundance of baryons present in the fluid:

$$\begin{aligned} w = \frac{dp}{d\rho} &= \frac{dp/dT}{d\rho/dT} \\ &= \frac{(dp_\gamma/dT) + (dp_b/dT)}{(d\rho_\gamma/dT) + (d\rho_b/dT)} \\ &= \frac{(4\rho_\gamma/3T) + 0}{(4\rho_\gamma/T) + (3\rho_b/T)} \\ &= \frac{4/3}{4 + (3\Omega_b a/\Omega_\gamma)} \\ &\approx \frac{1}{3(1 + 680\, a)}. \end{aligned} \tag{7.20}$$

The subscripts γ and b in these expressions denote photons and baryons, respectively, and in the last step I have used the measured values of Ω_γ and Ω_b. From this, we see that the sound speed transitioned from $c_s \approx 0.58c$ in the very early universe, to $c_s \approx 0.53c$ at the time of matter-radiation equality, and to $c_s \approx 0.45c$ at the time of last scattering. Inserting this result into Eq. 7.19, one finds that the sound horizon at the time of last scattering was $d_s \approx 0.126$ Mpc.

The sound horizon sets not only the scale of the first acoustic peak, $k_1 = \pi/d_s$, but also those of the higher peaks. Because the frequency of a given mode is proportional to its wavenumber, all modes that are integer multiples of k_1 will be at an extremum at the time of recombination, $k_2 = 2k_1$, $k_3 = 3k_1$, etc. The odd-numbered peaks are the result of those modes that were in a state of maximum compression at recombination, while the even-numbered peaks correspond to those that were in a state of maximum rarefaction. As we will see in Sec. 7.6, the locations and heights of these peaks tell us a great deal about our universe, including its geometry and composition.

7.5.4 The damping tail

Through diffusion, photons can travel from high density to low density regions, washing out the features that would have otherwise been seen in the CMB's angular power spectrum on very small scales. This effect is referred to as diffusion damping, or Silk damping [37].

To estimate how diffusion will impact the angular power spectrum of the CMB, consider that over a random walk of N steps, each of a distance δd, a typical object will be displaced by a total distance of $d \sim \delta d \sqrt{N}$. In the application at hand, the step size is the mean free path for Thomson scattering, $\delta d \approx (n_e \sigma_t)^{-1}$, while the number of steps is equal to the total time multiplied by the scattering rate, $N \approx t\,(n_e \sigma_t c)$. Taking the derivative of the diffusion length with respect to time, we can determine the effective net velocity of this random motion:

$$\begin{aligned} v_{\rm dif} &\sim \frac{d}{dt}(\delta d \sqrt{N}) \\ &\approx \frac{d}{dt}\left(\frac{ct}{n_e \sigma_t}\right)^{1/2} \\ &\approx \frac{1}{2}\left(\frac{c}{t\, n_e \sigma_t}\right)^{1/2}, \end{aligned} \tag{7.21}$$

where $n_e = n_{b,0}\,(1 - Y_p/2)\,a^{-3}\,X_e$ is the number density of free electrons (see Secs. 6.1 and 6.2). Over the history of our universe up to the time of last scattering, the distance that a typical photon is displaced via diffusion can be estimated by Eq. 7.19, but using the diffusion velocity in place of c_s:

$$\begin{aligned} d_{\rm dif} &= a_{\rm ls} \int_0^{a_{\rm ls}} \frac{v_{\rm dif}(a)\, da}{H(a)\, a^2} \\ &= \frac{a_{\rm ls}}{2\sigma_t^{1/2}} \int_0^{a_{\rm ls}} \frac{da}{H(a)\, a^2} \left(\frac{1}{t(a)\, n_e(a)}\right)^{1/2} \\ &= \frac{a_{\rm ls}}{2\sigma_t^{1/2}\, n_{b,0}^{1/2}\, (1 - Y_p/2)^{1/2}} \int_0^{a_{\rm ls}} \frac{da}{H(a)\, a^{1/2}\, t(a)^{1/2}\, X_e(a)^{1/2}}. \end{aligned} \tag{7.22}$$

Performing this integral for $X_e = 1$, we find that diffusion will wash out any power on length scales smaller than $d_{\rm dif} \lesssim 8\,{\rm kpc}$, corresponding to angular scales of $l \gtrsim l_1 \times (d_s/d_{\rm dif}) \sim 3{,}500$, where $l_1 \sim 220$ is the location of the first peak. Including the evolution of X_e in this calculation, as discussed in the previous chapter, we find that a significant contribution to $d_{\rm dif}$ comes from the diffusion that took place around the time of last scattering, when the universe was no longer fully ionized (i.e., X_e was no longer near unity). When this is properly taken into account, one finds that the angular power spectrum of the CMB is suppressed by diffusion on all scales $l \gtrsim 1{,}000{-}2{,}000$.

The motion of neutrinos around the time of last scattering also suppresses the small scale power of the CMB. Unlike photons, neutrinos are entirely decoupled well before the time of last scattering, allowing them to freely travel from high density to low density regions. The impact of the free streaming of a decoupled particle species on the CMB's angular power spectrum depends on the velocities of the particles, which in turn depend on their mass and temperature. Since they are very light, $m_\nu \lesssim 0.1$ eV, Standard Model neutrinos will free stream much more rapidly than photons can diffuse, thereby working to erase density

perturbations that would have otherwise existed on small angular scales. Non-relativistic decoupled particles (if they exist), would free stream more slowly, limiting their impact to smaller angular scales [38].

7.5.5 Secondary effects

The most prominent features that are observed in the CMB's angular power spectrum were embedded into it by or around the time of last scattering. As these photons subsequently traveled through the universe, however, their properties were also subtly impacted by the scattering and gravitational deflection that took place, especially at low redshifts. These kinds of phenomena are referred to as secondary effects, some of the most important of which include the following:

- Isotropic scattering: As mentioned in Chapter 6, the scattering of photons with free electrons during their propagation since recombination suppresses the acoustic peaks in the angular power spectrum by a factor of $e^{2\tau}$, where τ is the optical depth as integrated from time of last scattering to the present. This suppression is approximately scale-invariant and simply lowers all of the acoustic peaks in concert.
- Gravitational lensing: As the photons that make up the CMB pass by or through dense structures in the low-redshift universe, they can be deflected, washing out small scale power and suppressing the acoustic peaks at high *l*s.
- The Sunyaev-Zel'dovich effect: As CMB photons pass through galaxy clusters, they can undergo Compton scattering with high-energy electrons, increasing their frequencies. Like other secondary effects, this largely impacts the CMB at high *l*s. The Sunyaev-Zel'dovich effect also provides a useful way for cosmologists to discover distant galaxy clusters [39].

7.6 Extracting cosmological parameters

7.6.1 The geometry of our universe

In Sec. 7.5.3, we calculated that the largest regions that were in a state of maximum compression at the time of last scattering were $d_s \approx 0.126$ Mpc in extent, corresponding to the size of the sound horizon. The size of these regions determines, in part, the location of the first peak in the CMB's angular power spectrum. The location of this peak, however, also depends on the geometry of the space that these photons have traversed since they began free streaming.

Consider a region located on the surface of last scattering with a size given by $d_s \approx 0.126$ Mpc. This region, as it appears in the CMB today, will subtend an angle given by the following:

$$\theta_s = \frac{d_s}{d_A(z_{\text{ls}})} = \begin{cases} \dfrac{d_s(1+z_{\text{ls}})H_0\sqrt{|\Omega_\kappa|}}{c\,\sin(\sqrt{|\Omega_\kappa|}d_C H_0/c)}, & \kappa > 0, \\ \dfrac{d_s(1+z_{\text{ls}})}{d_C}, & \kappa = 0, \\ \dfrac{d_s(1+z_{\text{ls}})H_0\sqrt{\Omega_\kappa}}{c\,\sinh(\sqrt{\Omega_\kappa}d_C H_0/c)}, & \kappa < 0, \end{cases} \tag{7.23}$$

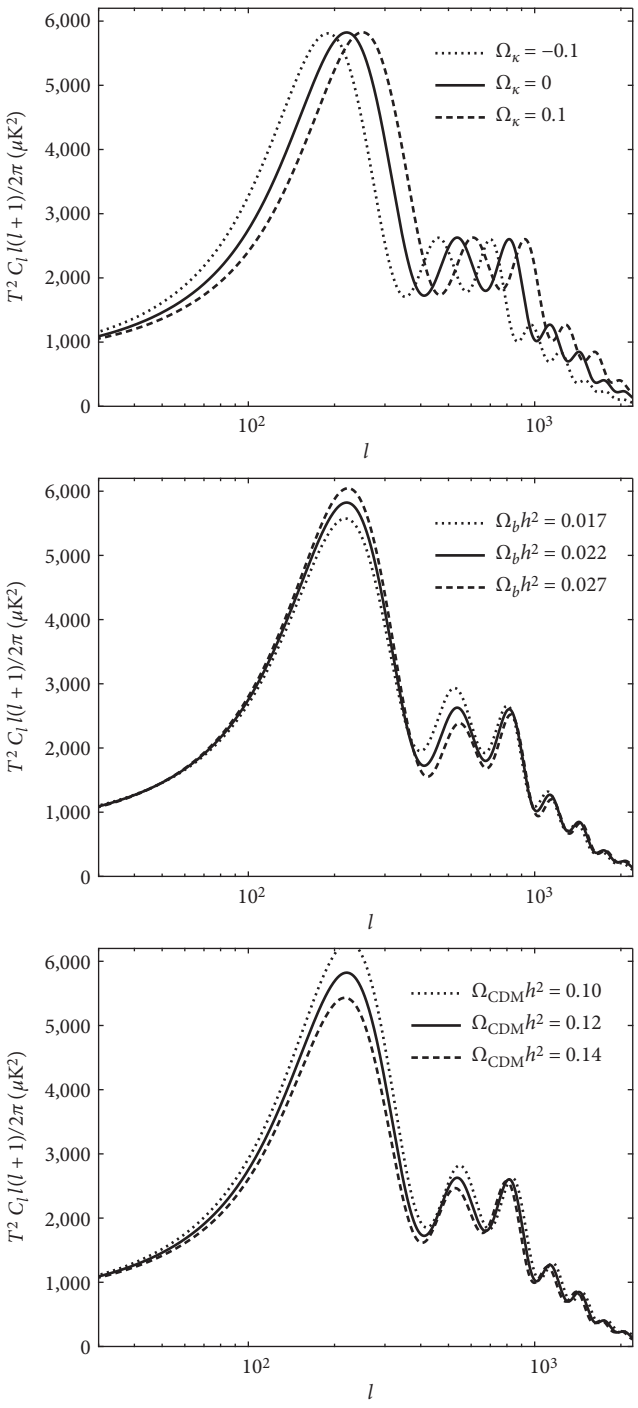


Figure 7.4 The temperature angular power spectrum of the CMB for various choices of the spatial curvature, Ω_κ, the baryon abundance, $\Omega_b h^2$, and the abundance of (cold) dark matter, $\Omega_{\rm CDM} h^2$, where $h \equiv H_0/(100\,{\rm km/s/Mpc})$. The other parameters have all been set to their best-fit values, adjusting Ω_Λ to obtain each value of Ω_κ (top), or to maintain $\Omega_{\rm tot} = 1$ (other frames). These curves were calculated using the publicly available code CAMB [40].

where d_A evaluated at $z_{\rm ls} = 1090$ is the angular diameter distance to the surface of last scattering (see Eq. 2.59), and d_C is the comoving distance to $z = 1090$ (Eq. 2.57). Importantly, the angular extent of these regions depends on the curvature of our universe, $\Omega_\kappa \equiv 1 - \Omega_{\rm tot}$ (recall that $\kappa > 0$ if $\Omega_\kappa < 0$ and $\kappa < 0$ if $\Omega_\kappa > 0$). This is because in a universe with $\kappa > 0$, lines will converge over long distances, causing distant objects to subtend a larger angle than they would have in Euclidean space. Similarly, such objects would subtend a smaller angle if viewed through a space with negative curvature.

Taking a Taylor series expansion of either sin and sinh in the above expressions, we find the following expression for θ_s in the small curvature limit:

$$\theta_s \approx \frac{d_s(1+z_{\rm ls})}{d_C}\left[\frac{1}{1+\Omega_\kappa d_C^2 H_0^2/6c^2}\right], \qquad |\Omega_\kappa| \ll (c\pi/H_0 d_C)^2. \tag{7.24}$$

By precisely measuring the locations of the acoustic peaks in the CMB's angular power spectrum, one can determine the value of Ω_κ. Measurements by the Planck satellite have found consistency with $\Omega_\kappa = 0$, and can rule out models in which $\Omega_{\rm tot}$ differs from unity by more than a fraction of a percent.

7.6.2 The baryon abundance

In the previous subsection, I showed that by measuring the locations of the acoustic peaks in the CMB's angular power spectrum, one could determine the overall geometry of our universe. The locations of these peaks, however, also depend on the sound speed, which in turn depends on the abundance of baryons (see Eq. 7.20). If all we had to go on were the locations of the acoustic peaks, the degeneracy between the effects of Ω_b and Ω_κ would represent a serious obstacle for determining these and other cosmological parameters. Fortunately, the baryon density not only impacts the locations of the acoustic peaks, but also the magnitudes of these peaks, providing us with a way to independently constrain the values of these parameters, as well as that of the total matter abundance, Ω_M.

The impact of the baryon density is most directly seen in the ratio of the heights of the even- and odd-numbered peaks in the CMB's angular power spectrum. Recall that the odd-numbered peaks correspond to regions that were in a state of maximum compression at the time of last scattering, whereas the even-numbered peaks are associated with regions that were in a state of maximum rarefaction. As these acoustic waves oscillate, the baryons (being nonrelativistic) act as mass, increasing the degree to which the regions compress at their extrema, but not impacting the degree to which they expand. Thus the ratio of the height of the first peak to that of the second peak, for example, provides us with a direct measurement of the abundance of baryons in the early universe.

In summary, the abundance of baryons impacts the angular power spectrum of the CMB in at least three significant ways: 1) by setting the sound speed and thus impacting the locations of the acoustic peaks, 2) by determining the maximum degree of compression and thus setting the ratio of the heights of the odd- and even-numbered acoustic peaks, and 3) by impacting the rate of diffusion and the corresponding suppression of power on small angular scales.

7.6.3 The dark matter abundance

Most of the matter in our universe is not made up of baryons, but of dark matter which was decoupled from the baryon-photon fluid during the formation and evolution of the acoustic oscillations. Unlike baryons, the dark matter did not participate in the acoustic oscillations

except through the effects of its gravity. The presence of dark matter has the overall effect of suppressing the amplitudes of the CMB's acoustic peaks on all scales, but especially at low multipoles.

During the era of radiation domination, $t \lesssim t_{\rm eq} \approx 51{,}000$ yr, the photons both dominated the energy density and provided the pressure that resisted compression. Whenever a given region reached a maximum state of compression during this era, there was a period in which the pressure support made it possible for the density fluctuation to decay as the space it occupied expanded. As a result, the baryon-photon fluid had less gravitational potential to work against as it rebounded, significantly increasing the amplitude of the acoustic oscillation. If it were not for dark matter, this enhancement would have occurred on all scales.

Once dark matter began to dominate the energy density of our universe, however, the source of the gravitational potentials (the dark matter) became distinct from the source of the pressure (the photons), putting an end to the decay of the gravitational potentials. As a result, only those modes that had already reached a state of maximum compression prior to matter domination were enhanced by this effect. For this reason, the first two peaks are most strongly suppressed by the presence of dark matter. By measuring the ratio of the heights of these first two peaks to those of the higher-l peaks, one can obtain a powerful measurement of the abundance of dark matter.

7.7 CMB polarization

So far in this chapter, I have focused entirely on the CMB's temperature. While certainly very important, this quantity does not encapsulate the entirety of the information contained within the CMB. In addition to the CMB's temperature, the polarization of the photons that make up the CMB can help us to constrain the cosmological parameters and otherwise probe the state and evolution of our universe.

The CMB became linearly polarized through the process of Thomson scattering. Before recombination, however, each photon scattered with many different electrons, washing out any degree of polarization that an observer might later hope to measure. In contrast, as the scattering rate fell below the Hubble rate at around the time of recombination, it became possible for the photons that constitute the CMB to become imbued with a lasting polarization signal. This occurred wherever a quadrupole anisotropy was present in the temperature field near the surface of last scattering. In other words, to create net polarization, the CMB had to have hot and cold regions that were separated by 90°, enabling Thomson scattering to leave the radiation in a linearly polarized state. The presence of polarization in the CMB thus provides us with an independent probe of the temperature anisotropies that were present at around the time of recombination, as our universe was becoming transparent to photons.

Similarly to its temperature distribution, measurements of the CMB's polarization can be used to constrain cosmological parameters, and provides an important consistency check with those determinations obtained through temperature information alone. In fact, the CMB's polarization carries more information than its temperature does, since the temperature in the given direction can be completely described in terms of a magnitude, while the polarization involves both a magnitude and a direction.

The polarization of the CMB is typically decomposed into curl-free "E-mode" patterns and divergence-free "B-modes." The former are created through anisotropic Thomson

scattering near the surface of last scattering, as described above. *B*-mode polarization, in contrast, cannot be directly produced in this way, but is instead generated through the gravitational lensing of those *E*-mode signals, in particular by structures present in the low-redshift universe, $z \lesssim 3$. More interestingly, *B*-modes could also be generated by gravitational waves that were produced during inflation (see Chapter 13). Primordial *B*-modes, which are yet to be detected, are widely considered to be one of the most promising probes of the era of inflation.

The total body of information contained in the temperature and polarization of the CMB can be described in terms of four different sets of C_ls. The C_ls that we have been discussing up to this point, and that are plotted in Figure 7.3, describe the correlations between the temperatures of the CMB at different points of the sky, separated by a given angle. These are sometimes referred to as C_l^{TT} (if no superscripts are given, C_l can be safely assumed to refer to C_l^{TT}). In addition, we can also consider the analogous correlation between the *E*-mode signals, C_l^{EE}, and the *B*-mode signals, C_l^{BB}. Furthermore, because the temperature and *E*-mode signals are correlated, additional information is contained in the correlation spectrum between the temperature and the *E*-mode signals, C_l^{TE}. In the absence of parity-violating interactions, we expect C_l^{TB} and C_l^{BE} to be zero for all l [41, 42].

In Figure 7.5, I have plotted the measured spectrum of the C_l^{TE}s and the C_l^{EE}s, and compared them to the predictions of the best-fit cosmological model. Like the temperature angular power spectrum, the C_l^{TE}s and the C_l^{EE}s contain a series of acoustic peaks which result from oscillations in the photon-baryon fluid. A careful observer will notice that the peaks in the C_l^{EE} spectrum are almost perfectly out of phase with those of the temperature angular power spectrum, occurring for those modes with $k = \pi/2d_s$, $3\pi/2d_s$, $5\pi/2d_s$, etc. As I've already discussed, the peaks in the C_l^{TT}s are found at scales that correspond to regions of maximum compression or rarefaction. In contrast, the C_l^{EE}s reflect the variations in the velocity of (and thus in the scattering that took place within) the photon-baryon fluid. A region in a state of maximum compression or rarefaction is also necessarily in a state of minimum velocity, and vice versa.

7.8 The current state of CMB measurements

The temperature angular power spectrum of the CMB, combined with the C_l^{TE} and C_l^{EE} spectra, provides us with a great deal of information regarding the composition and evolution of our universe. From this dataset, the Planck Collaboration has derived constraints on a variety of cosmological parameters, for a wide range of assumptions and models [8]. Some of the most important takeaways from these investigations include the following:

- **Our universe is indistinguishable from flat.** The total energy density of our universe is consistent with being equal to the critical density, $\Omega_\kappa \equiv 1 - \Omega_{\rm tot} = -0.0106 \pm 0.0065$. When combined with baryon acoustic oscillation (BAO) data (see Chapter 10), this determination tightens to $\Omega_\kappa = 0.0007 \pm 0.0019$. This result provides support for theories of inflation, which in most cases predict $\Omega_{\rm tot} = 1$ (see Chapter 13).
- **Our universe is dominated by dark energy and dark matter.** The single largest component of our universe's energy density takes the form of dark energy, with $\Omega_\Lambda = 0.6847 \pm 0.0073$. To the best of our ability to currently measure, this dark energy appears to evolve approximately as a cosmological constant (see

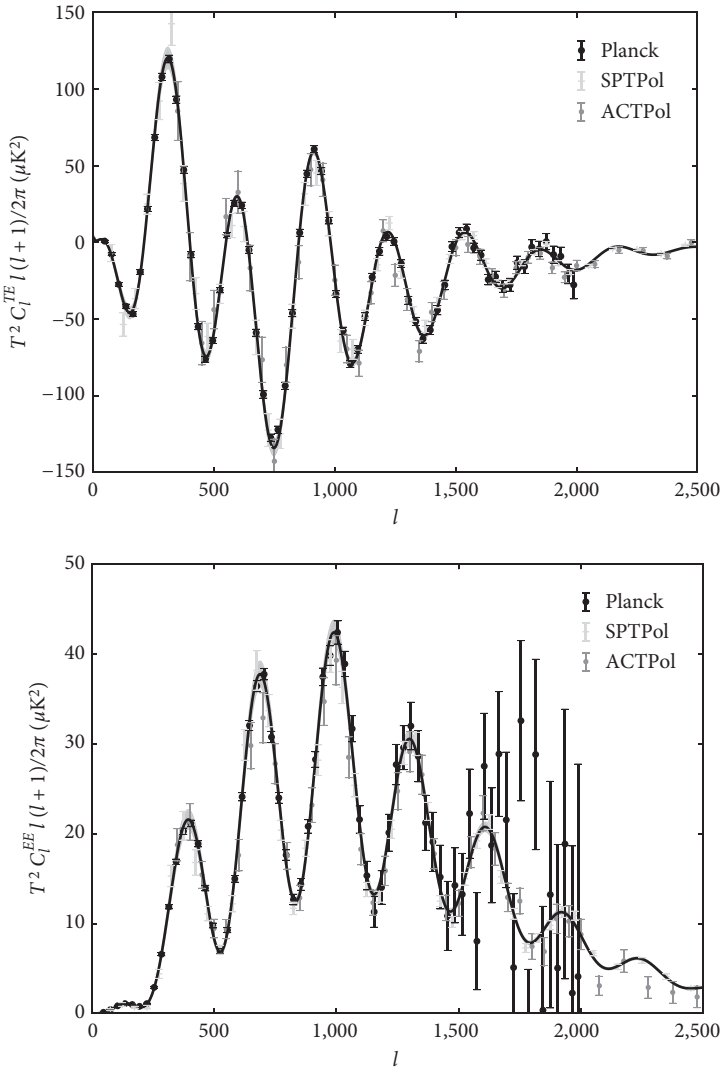


Figure 7.5 The CMB temperature anisotropy and *E*-mode polarization cross-power spectrum (top), and the angular power spectrum of *E*-mode polarization, as measured by the Planck satellite [35], SPTPol [43], and ACTPol [44, 45]. The solid curve in each frame is the prediction made by the best-fit ΛCDM cosmological model, and the grey bands around that curve represent the inherent uncertainty associated with cosmic variance.

Chapter 14). The remaining energy density is almost entirely made up of matter, $\Omega_M = 0.3153 \pm 0.0073$, of which only a small fraction is in the from of baryons: $\Omega_b h^2 = 0.02237 \pm 0.00015$, compared to $\Omega_{\rm CDM} h^2 = 0.1200 \pm 0.0012$ in cold dark matter.[1]

- **The primordial density perturbations were nearly scale-invariant.** The spectral index of the primordial density perturbations is $n_s = 0.9649 \pm 0.0042$, only slightly

[1] The quantity h is the current Hubble rate normalized such that $h \equiv H_0/(100\,\text{km/s/Mpc})$.

below the scale-invariant value of 1. Along with the measured flatness of our universe, this result provides substantial support for inflation (see Chapter 13).

- **The neutrinos are very light**. The masses of the three Standard Model neutrinos determine how much these particles free stream, and thus to what degree the small scale power in the CMB's angular power spectrum is washed out. Present CMB data have been used to constrain the sum of these three masses to be $\sum m_\nu < 0.24\,\text{eV}$. When combined with baryon acoustic oscillation (BAO) data, this constraint tightens to $\sum m_\nu < 0.12\,\text{eV}$. Note that this result is not very far from the *minimum* value of this quantity that is consistent with neutrino oscillation data, $\sum m_\nu > 0.06\,\text{eV}$ (see Chapter 11).

7.9 Dark radiation

In addition to providing us with powerful ways of measuring the parameters of the ΛCDM model, the information contained within the CMB can also be used to place constraints on various forms of exotic physics. In particular, the damping tail of the CMB is sensitive to the presence of feebly interacting relativistic particles, known as dark radiation.

The rate at which space was expanding in the early universe depends on the energy density, as dictated by the first Friedmann equation. In standard cosmology, the energy density during the first few hundred thousand years after the Big Bang was dominated by a mixture of matter (including both dark matter and baryons) and relativistic particles, namely photons and neutrinos (including antineutrinos). Using Eq. 4.6, the energy density of these relativistic particles can be written in terms of their temperatures:

$$\rho_\gamma = \frac{\pi^2 g_\gamma T^4}{30}, \qquad \rho_\nu \approx \frac{7}{8}\frac{\pi^2 g_\nu T_\nu^4}{30}, \tag{7.25}$$

where $g_\gamma = 2$ and $g_\nu = 6$ count the number of internal degrees of freedom of these particles, and the quantities T and T_ν are the temperatures of the photons and neutrinos, respectively (recall that the neutrinos decoupled from the photons roughly a second after the Big Bang). In Sec. 4.5, we found that for $T \ll m_e$, $T_\nu \approx (4/11)^{1/3}\,T$, allowing us to write the total energy density as follows:

$$\begin{aligned}\rho_{\text{tot}} &\approx \frac{\pi^2 g_\gamma T^4}{30} + \frac{7}{8}\frac{\pi^2 g_\nu T^4}{30}\left(\frac{4}{11}\right)^{4/3} \\ &\approx \rho_\gamma\left[1 + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\text{eff}}\right].\end{aligned} \tag{7.26}$$

In the last line, I have introduced the quantity N_{eff}, which is known as the effective number of neutrino species. If the neutrinos had instantly decoupled at the time of electron-positron annihilation, N_{eff} would have been equal to 3 (one for each neutrino flavor). If one carefully accounts for the non-equilibrium heating of the neutrinos during this period, it turns out that the final energy density of neutrinos is slightly higher, $N_{\text{eff}} \approx 3.046$.

In practice, cosmologists use the quantity N_{eff} to parameterize the energy density of radiation other than photons that was present during the era of recombination (or during BBN). If any exotic relativistic particles were present with a significant energy density during this period of time, they would have impacted the expansion rate of our universe, altering

the CMB's angular power spectrum. In order to have gone undetected, such particles would have to have been very feebly interacting. Many possible candidates for such "dark radiation" have been proposed, including axions, sterile neutrinos, gravitons, and hidden photons, to name a few.

It is convenient to parameterize the energy density in dark radiation in terms of its contribution to $N_{\rm eff}$ and the energy density in photons:

$$\rho_{\rm DR} = \Delta N_{\rm eff} \times \frac{7}{8}\left(\frac{4}{11}\right)^{4/3}\rho_\gamma, \tag{7.27}$$

where $\Delta N_{\rm eff} \equiv N_{\rm eff} - 3.046$.

Any relativistic particles that were present throughout the early universe will contribute to $N_{\rm eff}$, and by measuring this quantity and comparing the results to the Standard Model prediction of 3.046, we can constrain a wide variety of scenarios involving new light and feebly coupled particles. At present, CMB data yield a measurement of $N_{\rm eff} = 2.92^{+0.36}_{-0.37}$ (or $N_{\rm eff} = 2.99^{+0.34}_{-0.32}$ if BAO data is included) [8], consistent with the Standard Model prediction and leaving relatively little room for exotic forms of radiation.

To get a better idea of how restrictive this constraint is, consider a generic stable (or very long-lived) particle that at some point in the early universe was in thermal equilibrium with the Standard Model. If this particle species is light enough to remain relativistic throughout the era of recombination ($m \lesssim$ eV), it will act as dark radiation. At the temperature of this particle species' decoupling from the Standard Model bath, its energy density was given by $\rho_X = a\,\pi^2 g_X T_{\rm dc}^4/30$, where $a = 1$ if X is a boson or $a = 7/8$ if it is a fermion, and g_X is the number of internal degrees of freedom of the exotic particle. From entropy conservation, we can relate the subsequent temperature of the exotic species to the temperature of the photons:

$$T_X = T\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})}\right)^{1/3}. \tag{7.28}$$

Notice that this is very similar to the calculation of the neutrino temperature after its decoupling, as described in Sec. 4.5. From this, we can determine the energy density of the exotic species:

$$\begin{aligned}\rho_X &= \frac{a\,\pi^2 g_X T_X^4}{30} \\ &= \frac{a\,\pi^2 g_X T^4}{30}\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})}\right)^{4/3} \\ &= \frac{a\,g_X \rho_\gamma}{2}\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})}\right)^{4/3}.\end{aligned} \tag{7.29}$$

This energy density corresponds to the following contribution to $N_{\rm eff}$:

$$\Delta N_{{\rm eff},X} = \begin{cases} \dfrac{4\,g_X}{7}\left(\dfrac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})}\right)^{4/3}\left(\dfrac{11}{4}\right)^{4/3} & \text{(bosons)} \\ \dfrac{g_X}{2}\left(\dfrac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})}\right)^{4/3}\left(\dfrac{11}{4}\right)^{4/3} & \text{(fermions).} \end{cases} \tag{7.30}$$

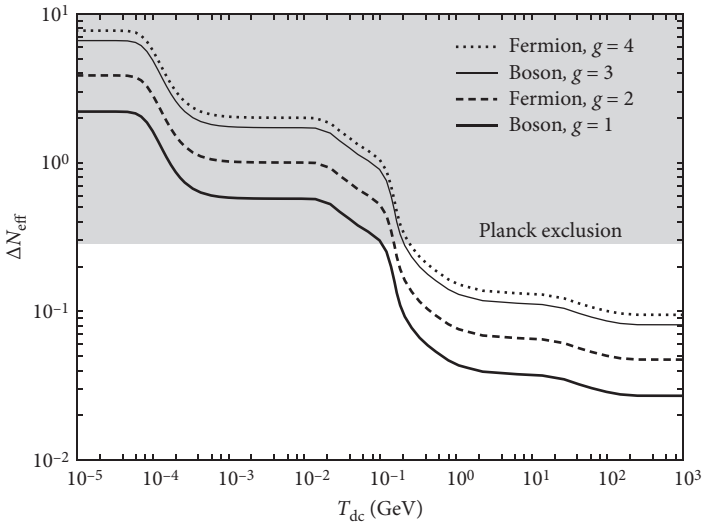


Figure 7.6 The contribution to the effective number of neutrino species from a stable, relativistic particle species, as a function of the temperature at which it decoupled from the Standard Model. CMB measurements rule out any contributions larger than $\Delta N_{\rm eff} > 0.29$ (the grey-shaded region).

Consider, for example, the impact of a light Dirac fermion ($a = 7/8$, $g_X = 4$) that decoupled from the Standard Model at a very high temperature, $T_{\rm dc} \gtrsim$ TeV, corresponding to $g_{\star,S}(T_{\rm dc}) \approx 106.75$. In this case, the contribution of this particle species to $N_{\rm eff}$ would be $\Delta N_{{\rm eff},X} = 2\,(3.909/106.75)^{4/3}(11/4)^{4/3} \approx 0.094$, which is consistent with current measurements of the CMB. If this particle had decoupled later, the entropy dumps within the Standard Model bath would have been more modest (i.e., the value of $g_{\star,S}$ at decoupling would have been smaller), resulting in a larger contribution to $N_{\rm eff}$. At present, measurements of the CMB rule out any light, stable particle species that decoupled from the Standard Model bath at a temperature below a few hundred MeV (around the time of the QCD phase transition). Future CMB measurements are projected to be sensitive to values of $\Delta N_{\rm eff}$ as small as ~ 0.02–0.03 [46], dramatically increasing our ability to detect the effects of light thermal relics.

Additional reading

There are a number of textbooks that cover the evolution of density perturbations in the early universe and their impact on the CMB in considerably more detail than I have done here. If you are looking to dig deeper into this subject, I recommend Dodelson's *Modern Cosmology*, Baumann's *Cosmology*, Weinberg's *Cosmology*, and Gorbunov and Rubakov's *Theory of the Early Universe*.

Problems

7.1 Show that the Planck blackbody spectrum of Eq. 7.2 is equivalent to the equilibrium distribution given in Eq. 7.1.

7.2 Compare the measurements of the CMB's spectrum to that of the sun. How much do these spectra each depart from that of an idealized blackbody.

7.3 At what redshift was the temperature of the CMB equal to the boiling point of water? At that redshift, what was the approximate age of our universe, and what was the energy density of the CMB?

7.4 At what temperature and time did the energy density in baryons exceed that in photons for the first time?

7.5 Assuming that the temperature variations of the CMB are indeed Gaussian, use the measured values of $\langle T \rangle$ and $\langle \delta T^2 \rangle^{1/2}$ to estimate the probability that a randomly selected region of the sky will feature a temperature greater than 2.7256 K or less than 2.7250 K.

7.6 Assuming that the temperature variations of the CMB are perfectly Gaussian, estimate the distance to the closest CMB photon with an energy greater than 10^{-3} eV.

7.7 Show that the expression for a_{lm} given in Eq. 7.9 follows from their definition in Eq. 7.7. Note that since the spherical harmonics make up an orthogonal and unit-normalized basis, $\int Y_{lm}(\hat{n})\, Y^{\star}_{l'm'}(\hat{n})\, d\hat{n} = \delta_{ll'}\, \delta_{mm'}$.

7.8 How much uncertainty in the value of C_l arises from cosmic variance for a choice of l that corresponds to 1° regions on the sky? What about 0.01° regions?

7.9 For the specific cases of $l = 0, 1, 2$, show that $\sum_{m=-l}^{l} P_l^m(x) P_l^{-m}(1) = P_l(x)$.

7.10 Imagine a universe in which the Thomson cross section is 100 times smaller than it is in ours. Qualitatively, how would this have impacted the observed angular power spectrum of the CMB? Consider the impact of this change both on the time and redshift of decoupling, and on the rate of diffusion.

7.11 Consider an object that is 1 kilometer in diameter and located at earth's equator. If light traveled in straight lines along the surface of the earth, what angle would this object subtend when viewed from the North Pole? What angle would it subtend if viewed from the same distance in Euclidean space?

7.12 Consider a region of space that is 1 Mpc in radius and located at a redshift of $z = 10$. Calculate the angular size of that region as it appears to us, assuming that we are living in a matter-only universe with $\Omega_M = 0.5$, $\Omega_M = 1$, or $\Omega_M = 1.5$.

7.13 Imagine a universe in which the total matter density consisted of baryonic matter. How would this have impacted the observed angular power spectrum of the CMB? Consider the impact of this change on the sound speed, on the ratio of the heights of the even and odd peaks, and on the rate of diffusion. In your reasoning and calculations, neglect any impact that the larger baryon abundance would have had on the ionization history or on the process of photon decoupling.

7.14 Consider a light and long-lived Majorana fermion ($g_X = 2$) that decoupled from the Standard Model bath when the temperature was $T = 10\,\text{TeV}$. Calculate the contribution from this particle species to N_{eff}. How would this result change if there existed several additional short-lived particle species in the thermal bath with masses near the TeV scale?

8

Big Bang Nucleosynthesis

The CMB provides us with a detailed snapshot of the era of recombination. To learn about the history and evolution of our universe at earlier times, we are forced to rely on other information. The nuclei of most of the light elements found in our universe today were forged through nuclear fusion during the first seconds and minutes after the Big Bang. The fact that the measured primordial abundances of these elements agree with the predictions of the Big Big theory gives us great confidence that our universe evolved more or less as expected from the first few seconds onward, without the presence of any significant quantities of unknown forms of energy or other exotic features (other than dark matter and dark energy). In this chapter, I describe the calculations behind this process, known as Big Bang nucleosynthesis (BBN), and discuss what measurements of the primordial element abundances can tell us about our universe's earliest measured era.

In the middle of the twentieth century, the Big Bang theory was competing for preeminence with cosmological models in which the universe was in a steady state configuration. In steady state models, the universe was expanding, but retained an approximately constant density through the spontaneous production of new matter. The advocates of these two cosmological paradigms would come to have very different views about the origin of nuclei. The steady state proponents argued that all of the nuclei found in our universe had been forged through nuclear fusion in stars. In contrast, George Gamow and Ralph Alpher proposed in a 1948 paper that nearly all of the nuclei had instead been created in the hot conditions that existed shortly after the Big Bang.[1] Both of these groups, it turns out, were wrong.

The calculations originally presented by Gamow and Alpher had a number of technical problems. When these mistakes were addressed, it became clear that although large amounts of deuterium and helium nuclei could have been generated in the early universe, along with a smattering of lithium, this era could not have been responsible for any of the heavier elements. For this reason, stellar nucleosynthesis remained the leading theory until the late 1950s. By that point, however, it had started to become evident that stars could not have produced nearly as much helium as is found in our universe. The final resolution to

[1] The paper by Gamow and Alpher that introduced BBN also included Hans Bethe on its author list. It seems that Bethe had nothing to do with the paper, but was included as part of a brilliant marketing strategy, allowing it to become known as the "Alpha-Beta-Gamma" paper.

this conundrum was a hybrid of these competing proposals. Most of the light elements were produced primordially, while stellar nucleosynthesis has been responsible for the production of the heavier nuclei.

8.1 Proton and neutron freeze out

During BBN, processes involving the strong and weak forces each played important roles. Because of the very different strengths of these two forces, however, their effects were most impactful during different periods of time. In particular, the weak force was active in transforming protons into neutrons and vice versa until about a second after the Big Bang, but was largely ineffective at later times (with the exception of being responsible for neutron decay). The strong force, in contrast, didn't begin to efficiently bind nucleons into nuclei until later, and played its most important role over a period of several minutes.

During radiation domination, the rate of Hubble expansion can be written as

$$\begin{aligned} H &= \left(\frac{8\pi}{3M_{\rm Pl}^2}\frac{\pi^2 g_\star(T) T^4}{30}\right)^{1/2} \\ &\approx 0.38\,{\rm s}^{-1} \times \left(\frac{g_\star(T)}{3.36}\right)^{1/2} \left(\frac{T}{\rm MeV}\right)^2. \end{aligned} \tag{8.1}$$

Just as we saw in Sec. 4.5 in the context of neutrino decoupling, the scattering rate of a given particle interacting through the weak force exceeded the rate of Hubble expansion only when $T \gtrsim$ MeV. As a result, once the temperature of our universe had fallen below this value, weak scattering processes were no longer capable of efficiently transferring energy between particles, or converting particles into other species. In contrast, the much larger cross sections associated with the strong force led to scattering rates that exceeded that of Hubble expansion up to much later times.

In this section, we will focus on the period of time leading up to the freeze out of the weak interactions, at which point no significant number of nuclei had yet formed. Since only the weak force can convert protons into neutrons (or vice versa), the timing of this transition sets the ratio of neutrons to protons in our universe, which is essential for determining how much helium and other nuclei would ultimately be produced through the process of BBN.

During this era, protons and neutrons were each present in significant abundances, and could transform into each other through charged-current weak interactions:

$$\begin{aligned} n + \nu_e &\longleftrightarrow p + e^-, \\ n + e^+ &\longleftrightarrow p + \bar{\nu}_e, \\ n &\longleftrightarrow p + e^- + \bar{\nu}_e. \end{aligned} \tag{8.2}$$

In equilibrium, it follows from Eq. 4.8 that the neutron-to-proton ratio will be given by

$$\begin{aligned} \frac{n_n}{n_p} &= \frac{g_n}{g_p}\left(\frac{m_n}{m_p}\right)^{3/2} e^{-(m_n - m_p + \mu_n - \mu_p)/T} \\ &\approx e^{-Q/T}\, e^{(\mu_n - \mu_p)/T}, \end{aligned} \tag{8.3}$$

where $Q \equiv m_n - m_p \approx 1.293$ MeV. In chemical equilibrium, each of the processes in Eq. 8.2 enforces the following relationship between the chemical potentials: $\mu_n + \mu_{\nu_e} = \mu_p + \mu_e$. In practice, however, the chemical potentials are negligible in this application, and the equilibrium ratio is simply given by $n_n/n_p \approx e^{-Q/T}$.

As in the case of recombination, the relevant scattering rates are rapid enough at high temperatures to keep the neutron-to-proton ratio near its equilibrium value. This does not remain true, however, as the temperature continues to drop. To evaluate when this transition occurs, we need to consider the relevant cross sections and number densities. In the era of neutron-proton freeze out, the 2-to-2 processes are most important. In particular, the cross section for the process $n + e^+ \to p + \bar{\nu}_e$ is given by

$$\sigma = \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2 E_\nu^2}{32\pi m_W^4 v_e}, \tag{8.4}$$

where g_2 is the $SU(2)$ gauge coupling, $g_A \approx 1.26$ is the measured value for the axial-vector coupling of beta decay, $V_{ud} \approx 0.974$ is the appropriate entry of the CKM matrix, E_ν is the energy of the outgoing antineutrino, m_W is the mass of the W boson, and v_e is the velocity of the incoming positron.

Combining this with the equilibrium number density of positrons from Eq. 4.5, the rate for this process per neutron is given by

$$\begin{aligned} \Gamma_{n+e^+ \to p+\bar{\nu}_e} &\approx n_{e^+} \sigma v \\ &\approx \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2}{64\pi^3 m_W^4} \int_{m_e}^{\infty} \frac{(E_e^2 - m_e^2)^{1/2} E_e E_\nu^2 dE_e}{(e^{E_e/T} + 1)}. \end{aligned} \tag{8.5}$$

Relating $E_\nu = m_n - m_p + E_e = Q + E_e$, and including an additional factor of $1 - f_\nu$ to account for Pauli blocking (see Eqs. 4.29 and 4.30), we arrive at the following rate:

$$\begin{aligned} \Gamma_{n+e^+ \to p+\bar{\nu}_e} &= \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2}{64\pi^3 m_W^4} \int_{m_e}^{\infty} \frac{(E_e^2 - m_e^2)^{1/2} E_e (Q + E_e)^2 dE_e}{(e^{E_e/T} + 1)} \left[1 - \frac{1}{e^{(Q+E_e)/T} + 1} \right] \\ &= \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2}{64\pi^3 m_W^4} \int_{m_e}^{\infty} \frac{(E_e^2 - m_e^2)^{1/2} E_e (Q + E_e)^2 dE_e}{(e^{E_e/T} + 1)(e^{-(Q+E_e)/T} + 1)}. \end{aligned} \tag{8.6}$$

Repeating this exercise for the other neutron-to-proton converting scattering processes, we find

$$\begin{aligned} \Gamma_{n+\nu_e \to p+e^-} &= \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2}{64\pi^3 m_W^4} \int_0^{\infty} \frac{E_\nu^2 (Q + E_\nu)^2 dE_\nu}{(e^{E_\nu/T} + 1)} \left[1 - \frac{1}{e^{(Q+E_\nu)/T} + 1} \right] \\ &= \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2}{64\pi^3 m_W^4} \int_0^{\infty} \frac{E_\nu^2 (Q + E_\nu)^2 dE_\nu}{(e^{E_\nu/T} + 1)(e^{-(Q+E_\nu)/T} + 1)}, \end{aligned} \tag{8.7}$$

$$\begin{aligned} \Gamma_{p+e^- \to n+\nu_e} &= \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2}{64\pi^3 m_W^4} \int_{m_e}^{\infty} \frac{(E_e^2 - m_e^2)^{1/2} E_e (-Q + E_e)^2 dE_e}{(e^{E_e/T} + 1)} \left[1 - \frac{1}{e^{(-Q+E_e)/T} + 1} \right] \\ &= \frac{g_2^4 (1 + 3g_A^2) V_{ud}^2}{64\pi^3 m_W^4} \int_{m_e}^{\infty} \frac{(E_e^2 - m_e^2)^{1/2} E_e (-Q + E_e)^2 dE_e}{(e^{E_e/T} + 1)(e^{(Q-E_e)/T} + 1)}, \end{aligned} \tag{8.8}$$

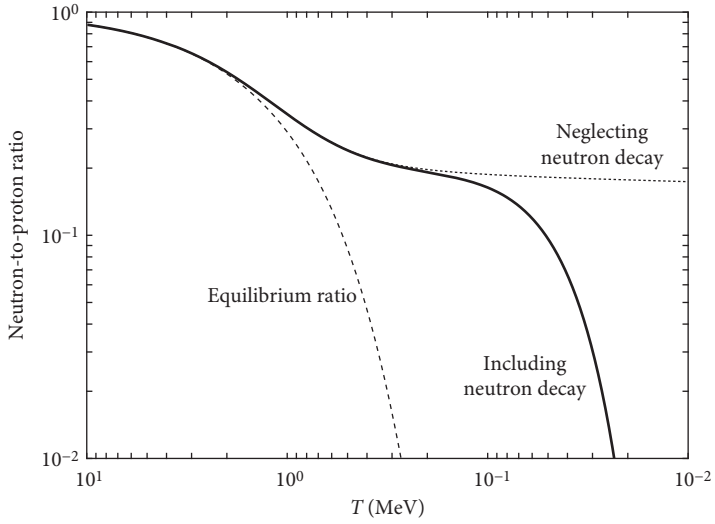


Figure 8.1 The neutron-to-proton ratio during the freeze out of the weak interactions. At $T \gg$ MeV, this ratio was held very close to its equilibrium value. This remained the case until the rate of proton-to-neutron conversion (and vice versa) fell below that of Hubble expansion. If it were not for neutron decay, this process would have resulted in an asymptotic value of $n_n/n_p \sim 0.2$.

and

$$\Gamma_{p+\bar{\nu}_e \to n+e^+} = \frac{g_2^4(1+3g_A^2)V_{ud}^2}{64\pi^3 m_W^4} \int_0^\infty \frac{E_\nu^2(-Q+E_\nu)^2 dE_\nu}{(e^{E_\nu/T}+1)} \left[1 - \frac{1}{e^{(-Q+E_\nu)/T}+1}\right]$$

$$= \frac{g_2^4(1+3g_A^2)V_{ud}^2}{64\pi^3 m_W^4} \int_0^\infty \frac{E_\nu^2(-Q+E_\nu)^2 dE_\nu}{(e^{E_\nu/T}+1)(e^{(Q-E_\nu)/T}+1)}. \tag{8.9}$$

While the process of neutron decay, $n \to p + e^- + \bar{\nu}_e$, is too slow to be important at the time of neutron-proton freeze out, it becomes relevant later, with important implications for BBN. The rate for neutron decay is simply given by

$$\Gamma_{n \to p+e^-+\bar{\nu}_e} = \frac{1}{\tau_n}, \tag{8.10}$$

where $\tau_n \approx 880.2$ s is the neutron lifetime.

Combining the rates for these processes with the effects of Hubble expansion, we arrive at the following Boltzmann equation for the number density of neutrons:

$$\frac{dn_n}{dt} = -3Hn_n + \Gamma_{p\to n} n_p - \Gamma_{n\to p} n_n \tag{8.11}$$

$$= -3Hn_n + \Gamma_{p\to n}[n_{b,0}a^{-3} - n_n] - \Gamma_{n\to p} n_n,$$

where $\Gamma_{p\to n}$ and $\Gamma_{n\to p}$ represent the sums of the relevant rates described above, and in the last line we have expressed the number density of protons in terms of those of baryons and neutrons, $n_p = n_b - n_n$.

In Figure 8.1, I show the numerical solution to Eq. 8.11, presented in terms of the neutron-to-proton ratio. At $T \gg$ MeV, this ratio is held very close to its equilibrium value.

This is maintained until the rates for the above-described scattering processes fall below that of Hubble expansion, at which point the neutron-to-proton ratio freezes out. This occurs at a temperature of $T \approx 0.8$ MeV. If it were not for neutron decay, this process would result in an asymptotic value of $n_n/n_p \sim 0.2$. Including neutron decay, but neglecting any formation of nuclei, this ratio evolves as follows after freeze out:

$$\frac{n_n}{n_p} \sim 0.2 \times e^{-t/\tau_n}. \tag{8.12}$$

In reality, the impact of neutron decay was limited by the fact that most of these neutrons became bound within helium nuclei before they could decay. Under the assumption that approximately all of the neutrons became bound into helium-4 nuclei shortly after freeze out, this result translates into a primordial helium mass fraction of $Y_p \equiv 4n_{\rm He}/(n_{\rm H} + 4n_{\rm He}) \approx 0.28$, which is not too much larger than the measured value of $Y_p \approx 0.245$. The difference between these two quantities reflects the fact that it took a nonnegligible fraction of τ_n for these neutrons to become bound within nuclei, somewhat limiting the amount of helium that was ultimately produced.

8.2 Nuclear abundances in equilibrium

The story of Big Bang nucleosynthesis began when our universe was at a temperature of several MeV. At this time, our universe was radiation-dominated and mostly populated with a hot bath of photons, electrons, positrons, neutrinos, and antineutrinos, along with a subdominant component of dark matter, and a relatively small number of protons and neutrons. By this time, there were very few antiprotons or antineutrons remaining in our universe, as the vast majority of these particles had already been destroyed through their annihilation with nucleons (see Chapter 12).

Much as we did for the process of recombination, it will be useful for us to first consider the abundances of the various nuclear species that would have been present if there had been adequate time to reach equilibrium. Once we have these equilibrium abundances in hand, we will consider the rates for various processes and evaluate the evolution of these abundances in a rapidly expanding universe.

In the temperature range under consideration, all nucleons and nuclei will be nonrelativistic, allowing us to write their equilibrium abundances as follows:

$$n_i = g_i \left(\frac{m_i T}{2\pi} \right)^{3/2} e^{(\mu_i - m_i)/T}, \tag{8.13}$$

where m_i, μ_i, and g_i are the mass, chemical potential, and the number of spin states of the species in question (see Table 8.1). Although the weak force can convert protons into neutrons and vice versa, the total number of baryons is conserved in all of the interactions that we will be concerned with here. This allows us to relate the chemical potentials according to $\mu_i = Z_i \mu_p + (A_i - Z_i)\mu_n$, where Z_i and A_i are the atomic number and atomic mass of the species, and μ_p and μ_n are the chemical potentials of the protons and neutrons, respectively. After some algebra, we can rewrite Eq. 8.13 as

$$n_i = g_i \, A_i^{3/2} \, 2^{-A_i} \left(\frac{2\pi}{m_{p,n} T} \right)^{3(A_i - 1)/2} n_p^{Z_i} \, n_n^{A_i - Z_i} \, e^{B_i/T}, \tag{8.14}$$

Table 8.1 The binding energies, spin, and number of spin degrees of freedom for various species of nuclei.

Isotope	B_i	Spin	g_i
^{1}H	–	1/2	2
^{2}H	2.22 MeV	1	3
^{3}H	8.48 MeV	1/2	2
^{3}He	7.72 MeV	1/2	2
^{4}He	28.3 MeV	0	1
^{6}Li	32.0 MeV	1	3
^{7}Li	39.2 MeV	3/2	4
^{9}Be	58.2 MeV	3/2	4
^{10}B	64.8 MeV	3	7
^{11}B	76.2 MeV	3/2	4
^{12}C	92.2 MeV	0	1
^{13}C	97.1 MeV	1/2	2
^{14}C	105.3 MeV	0	1

where n_p and n_n are the equilibrium number densities of protons and neutrons, and $B_i = Z_i m_p + (A_i - Z_i) m_n - m_i$ is the binding energy of species i (see Table 8.1).

It is useful in this context to consider the mass fraction of a given species:

$$X_i \equiv \frac{n_i \, A_i}{n_b}, \tag{8.15}$$

where $n_b = \eta \, n_\gamma$ is the total number density of baryons (regardless of whether bound or free), $\eta \approx 6 \times 10^{-10}$ is the measured baryon-to-photon ratio, and n_γ is the number density of photons. The equilibrium mass fraction of a nuclear species is given by

$$X_i = \frac{g_i \, A_i^{5/2} \, 2^{-A_i}}{n_b} \left(\frac{2\pi}{m_{p,n} T} \right)^{3(A_i - 1)/2} [X_p n_b]^{Z_i} \, [X_n n_b]^{A_i - Z_i} \, e^{B_i/T} \tag{8.16}$$

$$= g_i \, A_i^{5/2} \, 2^{-A_i} \left(\frac{2\pi}{m_{p,n} T} \right)^{3(A_i - 1)/2} X_p^{Z_i} \, X_n^{A_i - Z_i} \left[\frac{\eta \, 2\zeta(3) \, T^3}{\pi^2} \right]^{A_i - 1} e^{B_i/T},$$

where in the second line I have used $n_b = \eta \, n_\gamma = \eta \, 2\zeta(3) \, T^3/\pi^2$.

To determine the equilibrium abundance of a given species, X_i, we will need both Eq. 8.16, and the number densities of protons and neutrons. To this end, we can make use of the expression $n_b = \eta \, n_\gamma$, as well as the measured value of η, and adopt a neutron-to-proton ratio of $n_n/n_p = 0.14$ (I have chosen a slightly smaller value than the aforementioned 0.2 in order to account for the effects of neutron decay). This leaves us with a set of equations with a unique solution for the equilibrium abundances. In Figure 8.2, I show the numerical results for this set of equations, including all nuclear species with $Z \leq 6$.

What Figure 8.2 demonstrates is that if the universe had expanded arbitrarily slowly, allowing all nuclear processes adequate time to reach equilibrium, then the vast majority of the protons and neutrons would first become bound into helium-4 nuclei, and later into carbon-12. These two isotopes each stand out in part because of their high binding energies per nucleon. Neglecting the impact of the binding energies, Eq. 8.16 strongly favors light

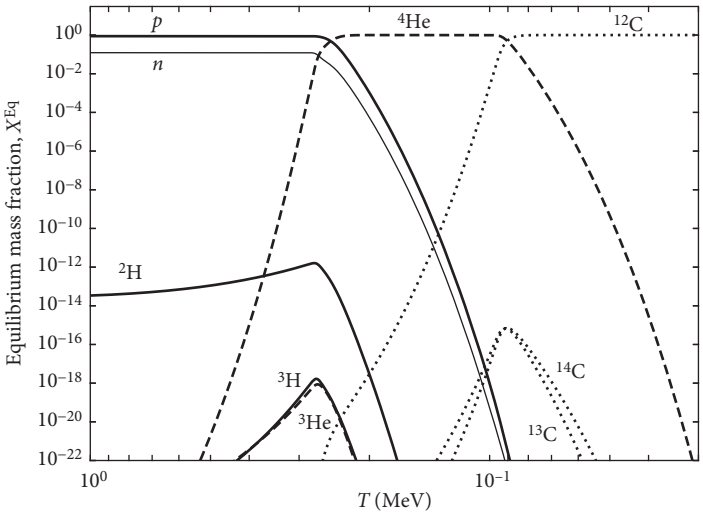


Figure 8.2 The equilibrium mass fraction of all nuclear species with $Z \leq 6$, for a neutron-to-proton ratio of 0.14. If the universe had expanded arbitrarily slowly, allowing all nuclear processes to reach and maintain equilibrium, the vast majority of the protons and neutrons would have become bound within helium-4 nuclei, and later within carbon-12.

isotopes, especially at low temperatures and when the abundance of free nucleons is small. The large binding energies of helium-4 and carbon-12, however, overcome these penalties in their respective temperature ranges.

Of course, primordial nucleosynthesis did not actually take place in a universe that was expanding at an arbitrarily slow rate. As a consequence, the nuclear processes at hand were often not able to maintain equilibrium, leading to qualitatively different outcomes than those suggested by Figure 8.2. In the following sections, we will explore the competition between the rates of the most important nuclear processes and the expansion of the universe during the formation of the light nuclei.

8.3 From nucleons to Helium

In this section, we will consider the three-way race between the binding of nucleons into nuclei, our universe's expansion rate, and the decay of neutrons. Once our universe had expanded and cooled enough to prevent fusion from efficiently taking place, the process of primordial nucleosynthesis came to an end. Similarly, no stable nuclei could form through the strong force without neutrons, forcing much of primordial nucleosynthesis to take place within the short window of time dictated by the neutron's lifetime.

The main processes at hand (see Figure 8.3) are the formation of deuterons,

$$p + n \rightarrow {}^{2}\mathrm{H} + \gamma, \tag{8.17}$$

followed by the formation of ^{3}H and ^{3}He nuclei,

$$\begin{aligned} {}^{2}\mathrm{H} + {}^{2}\mathrm{H} &\rightarrow {}^{3}\mathrm{H} + p, \\ {}^{2}\mathrm{H} + {}^{2}\mathrm{H} &\rightarrow {}^{3}\mathrm{He} + n, \end{aligned} \tag{8.18}$$

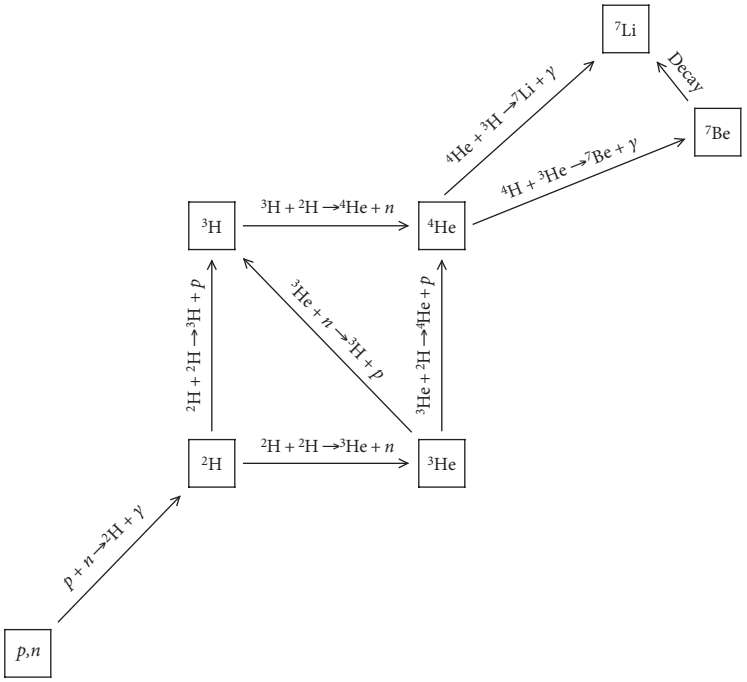


Figure 8.3 The most important nuclear processes in Big Bang nucleosynthesis.

and finally helium-4,

$$^2\mathrm{H} + ^3\mathrm{H} \rightarrow {}^4\mathrm{He} + n, \qquad (8.19)$$
$$^2\mathrm{H} + ^3\mathrm{He} \rightarrow {}^4\mathrm{He} + p.$$

The first step of deuteron formation is a relatively simple one. The cross section for this process ($\sigma v \sim 5 \times 10^{-20}\,\mathrm{cm}^3/\mathrm{s}$) and the number density of nucleons were both quite high, allowing the rate of deuteron formation per neutron to exceed that of Hubble expansion so long as $T \gtrsim 4 \times 10^{-4}\,\mathrm{MeV}$. As a result, the deuteron abundance was held very close to its equilibrium value throughout the period in which ^{3}H and ^{3}He were being formed. From Eq. 8.16, the equilibrium deuteron mass fraction is given by

$$X_{^2\mathrm{H}} \approx \frac{24\zeta(3)\eta}{\pi^{1/2}} X_p X_n \left(\frac{T}{m_p}\right)^{3/2} e^{B/T} \qquad (8.20)$$
$$\approx 3.4 \times 10^{-13} \times X_p X_n \left(\frac{T}{1\mathrm{MeV}}\right)^{3/2} e^{2.22\,\mathrm{MeV}/T}.$$

This relatively small deuteron abundance limited the rates at which ^{3}H and ^{3}He nuclei could be formed. For this reason, one might have guessed that processes such as $n + ^2\mathrm{H} \rightarrow {}^3\mathrm{H} + \gamma$ or $p + ^2\mathrm{H} \rightarrow {}^3\mathrm{He} + \gamma$ would have dominated over those involving the fusion of two deuterons. As it turns out, however, the cross sections for these processes are quite small, causing them to play only a minor role.

The cross sections for deuteron fusion are quite large, and vary significantly with energy. In the temperature range during which most of the ^{3}H and ^{3}He were produced

($T \sim 0.2 - 0.07$ MeV), both of the processes described in Eq. 8.18 have cross sections in the range of $\langle \sigma v \rangle \approx (1-5) \times 10^{-17}$ cm^3/s. The evolution of the abundances of these nuclei is described by the following Boltzmann equation:

$$\frac{dn}{dt} = -3Hn + \langle \sigma v \rangle \, n_{^2\mathrm{H}}^2, \tag{8.21}$$

where n denotes the combined number density of ^{3}H and ^{3}He. Using the equilibrium deuteron abundance given in Eq. 8.20, one finds that the rate of ^{3}H and ^{3}He production per deuteron increased as the temperature dropped, ultimately exceeding the rate of Hubble expansion once $T \lesssim 0.08$ MeV. This temperature corresponds to $t \approx 170$ s after the Big Bang, by which time approximately $1 - e^{-t/\tau_n} \approx 20\%$ of the neutrons that were present at the time of proton-neutron freeze out had already decayed. Once significant abundances of ^{3}H and ^{3}He had built up, these nuclei quickly fused with free deuterons to form helium-4.

Note that the timing of these events is quite sensitive to the overall number density of baryons. More specifically, if the baryon abundance had been larger, fusion would have started earlier, ultimately resulting in a larger fraction of neutrons that became bound into nuclei before they could decay. As a result, measurements of Y_p can be used to powerfully constrain the total abundance of baryons in our universe.

8.4 Beyond Helium

Up to this point, BBN had been a race to bind neutrons into deuterons before they could decay. Once significant numbers of deuterons were present, they quickly fused to form helium nuclei. The subsequent production of heavier nuclei, in contrast, was not limited by the lifetime of the neutron, but rather by the falling temperature, which suppressed the rates at which further fusion could occur.

There were two main factors that prevented the fusion of nuclei heavier than helium from proceeding efficiently. The first of these is the fact that there were never any particularly large quantities of deuterium, tritium, or helium-3 present in our universe, limiting the rates for processes involving these nuclear species. While a helium-4 nucleus could, in principle, fuse directly with a proton, this results in the production of ^{5}Li, which almost instantly decays with a lifetime of only $\sim 10^{-23}$ s. The formation of stable or relatively long-lived nuclei thus required the involvement of not only helium-4, but also deuterium, tritium, or helium-3.

At this point, you might ask about the fusion of two or more ^{4}He nuclei. In massive stars, with core temperatures above ~ 0.01 MeV, helium can fuse into carbon through the triple alpha process, $^4\mathrm{He} + {}^4\mathrm{He} \rightarrow {}^8\mathrm{Be}$, followed by ${}^8\mathrm{Be} + {}^4\mathrm{He} \rightarrow {}^{12}\mathrm{C} + 2\gamma$. The very short lifetime of ^{8}Be, however, makes this process very slow. While potentially effective over the lifetime of a star, this process did not have any appreciable impact when limited to the few minutes of BBN. More generally speaking, the lack of any stable nuclei with $A = 5 - 8$ significantly limited the avenues by which heavy nuclei could potentially be produced in the early universe.

The second major factor which limited the formation of heavy nuclei is the coulomb repulsion that occurs between nuclei with significant amounts of electric charge. The cross section for the fusion of two nuclei is proportional to the following:

$$\langle \sigma v \rangle \propto \exp\left(\frac{-1.9\,\bar{A}^{1/3} Z_1^{2/3} Z_2^{2/3}}{T_{\mathrm{MeV}}^{1/3}}\right), \tag{8.22}$$

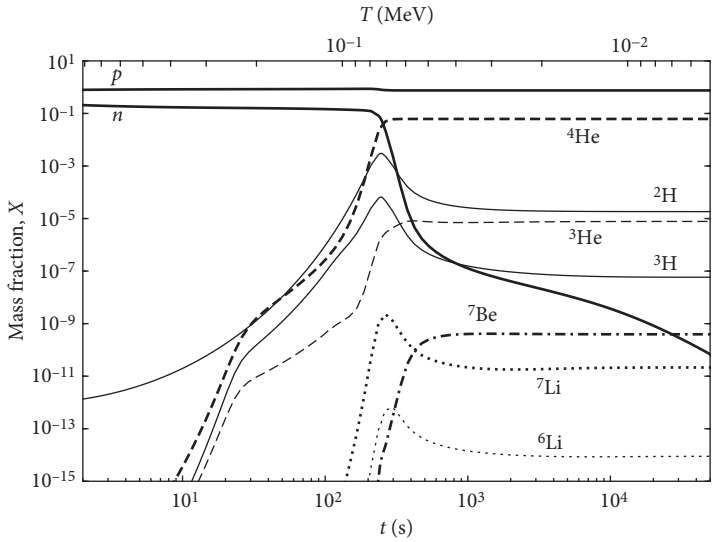


Figure 8.4 The evolution of the mass fractions of various nuclear species during Big Bang nucleosynthesis, as calculated using the publicly available code PRIMAT.

where $\bar{A} \equiv A_1 A_2/(A_1 + A_2)$, A_i and Z_i are the atomic mass and number of the incoming nuclei, and $T_{\rm MeV}$ is the temperature in units of MeV. As our universe expanded and its temperature dropped, this suppression became increasingly important, especially for nuclei containing multiple protons.

Despite these factors working against the formation of nuclei heavier than helium-4, small amounts of lithium and beryllium were produced through the following processes:

$$\begin{aligned} {}^4\mathrm{He} + {}^3\mathrm{He} &\rightarrow {}^7\mathrm{Be} + \gamma, \\ {}^4\mathrm{He} + {}^3\mathrm{H} &\rightarrow {}^7\mathrm{Li} + \gamma, \\ {}^4\mathrm{He} + {}^2\mathrm{H} &\rightarrow {}^6\mathrm{Li} + \gamma. \end{aligned} \tag{8.23}$$

Note that ^{7}Be decays through electron capture with a lifetime of 76.7 days, $^7\mathrm{Be} + e^- \rightarrow {}^7\mathrm{Li} + \nu_e$. As a result, the ultimate abundance of primordial ^{7}Li is equal to the sum of the abundances of ^{7}Be and ^{7}Li that were present prior to these decays. In effect, beryllium is the main source of primordial ^{7}Li.

As an example, consider the evolution of ^{7}Be, which is described reasonably well by the following Boltzmann equation:

$$\frac{dn_{\rm Be}}{dt} = -3Hn_{\rm Be} + \langle \sigma v \rangle \, n_{^4\mathrm{He}} \, n_{^3\mathrm{He}}. \tag{8.24}$$

The cross section in this expression is for the process $^4\mathrm{He} + {}^3\mathrm{He} \rightarrow {}^7\mathrm{Be} + \gamma$, and is approximately given by

$$\langle \sigma v \rangle \approx \frac{1.9 \times 10^{-18}\ \mathrm{cm^3/s}}{T_{\rm MeV}^{2/3}} \; e^{-5.7/T_{\rm MeV}}. \tag{8.25}$$

Using the number densities for ^{3}He and ^{4}He as shown in Figure 8.4, one finds that the production rate of beryllium modestly exceeded that of Hubble expansion during

our universe's first few hundred seconds, allowing the mass fraction of ^{7}Be to build up to roughly $\sim 10^{-9}$ over that time. In reality, the abundance of beryllium never got quite this large because its production rate was offset by processes which break it up, such as $^{7}\mathrm{Be} + n \rightarrow {}^{4}\mathrm{He} + {}^{4}\mathrm{He}$ and $^{7}\mathrm{Be} + {}^{2}\mathrm{H} \rightarrow p + {}^{4}\mathrm{He} + {}^{4}\mathrm{He}$.

8.5 The primordial element abundances

Predictions for the abundances of the various nuclear species that formed during BBN can be directly compared to measurements of the primordial nuclear abundances, providing us with a powerful probe of our universe's first seconds and minutes. Historically, such comparisons provided an important validation of the Big Bang theory. In modern decades, they have allowed us to further test the standard cosmological model, and to constrain its parameters.

Measurements of the primordial deuterium abundance are perhaps our single most powerful probe of the BBN era. A major challenge in these determinations, however, is the small binding energy of deuterium, which makes it easily destroyed in stars. Because of this, measurements of the ^{2}H/H ratio found in the interstellar medium, jovian planet atmospheres, and the solar wind each provides us with only a lower limit on the primordial value of this quantity. To determine how much deuterium was present in our universe prior to any processing in stars, we instead rely on observations of high-redshift (and thus very old) environments. In this regard, damped Lyman-alpha systems are particularly useful. Damped Lyman-alpha systems are dense regions in high-redshift galaxies, for which it is possible to use precision spectroscopy to measure the column densities of hydrogen and deuterium. At present, only a handful of such measurements have been made, but these have been used to provide a determination of $(^{2}\mathrm{H}/\mathrm{H})_p = (2.569 \pm 0.27) \times 10^{-5}$, where the subscript labels the fact that we are referring to the primordial value of this ratio. The measured value of this quantity doesn't show any correlation with redshift or metallicity (a measure of past star formation), supporting the conclusion that these measurements reflect the true primordial value. In contrast, measurements of the relative deuterium abundance in the solar system and in the Milky Way each show considerable variation, which is anticorrelated with the presence of metals, indicative of the impact of stellar processing.

The primordial helium mass fraction, Y_p, has been most effectively determined through measurements of helium and hydrogen emission lines from very low metallicity regions of partially ionized gas. These measurements reveal a correlation between the ^{4}He abundance and the presence of metals in the system, indicating that these regions have been contaminated by helium produced in stars. Extrapolating this relationship to zero metallicity results in a determination of $Y_p = 0.245 \pm 0.003$.

Determinations of the primordial abundance of helium-3 have proven difficult. Measurements of ^{3}He have been made in the solar system and the interstellar medium, each indicating $^{3}\mathrm{He}/\mathrm{H} = (1.1 \pm 0.2) \times 10^{-5}$. That being said, ^{3}He is both destroyed and created in stars at relevant rates, making it unclear whether these measurements are truly indicative of the primordial value. For this reason, most modern BBN analyses do not make use of any measurements of ^{3}He.

Lastly, I turn to lithium, whose primordial abundance is most readily determined by spectroscopic measurements of low-metallicity dwarf stars. Such measurements have been used to determine $(\mathrm{Li}/\mathrm{H})_p = (1.6 \pm 0.3) \times 10^{-10}$. Although this ratio represents the combined abundance of both ^{7}Li and ^{6}Li, the contribution from ^{6}Li is expected to be smaller

than that of ^{7}Li by a factor of $\sim 10^{-5}$. The observed relationship between this quantity and stellar metallicity has raised questions about whether the lithium abundance might be depleted in stars with moderate metallicities. If this turns out to be a significant effect, the measured value of Li/H should be treated as only an upper limit on the primordial value.

These measurements are plotted in Figure 8.5, along with the predicted primordial element abundances, as a function of either the baryon-to-photon ratio or (equivalently) the cosmological density of baryonic matter. By increasing the total abundance of baryons, we allow the fusion of deuterium to begin at a higher temperature, staring the whole process of BBN earlier. This results in a larger fraction of primordial neutrons ending up in ^{4}He, and leaving fewer behind to make up ^{2}H or ^{3}He. This trend can be easily identified in the upper and middle frames of Figure 8.5. Furthermore, we see that the measurements of Y_p and $(^2\mathrm{H/H})_p$ each favor a value of $\Omega_b h^2$ (or η) that is consistent with that determined from the angular power spectrum CMB, as measured by the Planck satellite (see Chapter 7). It is noteworthy that the favored range of $\Omega_b h^2$ corresponds to a density in baryons that is far less than the total density of matter, providing considerable support for the existence of non-baryonic dark matter.

The relationship between the baryon abundance and that of primordial ^{7}Li is more complicated. While the rate of direct lithium production, $^4\mathrm{He} + ^3\mathrm{H} \rightarrow ^7\mathrm{Li} + \gamma$, is reduced as one increases $\Omega_b h^2$, the amount that is produced through beryllium decays is increased. This results in the behavior seen in the lower frame of Figure 8.5. You will also notice from this figure that, for the range of $\Omega_b h^2$ favored by the CMB (and by the measured helium and deuterium abundances), the predicted abundance of ^{7}Li is $\sim$ 2–5 times larger than the measured value. As mentioned above, this might have to do with the depletion of lithium in stars. On the other hand, it could also be an indication that some kind of new physics is at work. At this point in time, the "lithium problem" of BBN remains unresolved.

8.6 BBN as a probe of exotic physics

In the past section, we saw that the predictions of standard BBN are in good agreement with observations (with the possible exception of the primordial lithium abundance). This gives us confidence that the thermal history of our universe played out approximately as we would have expected based on known physics, such as the kinds of particles that exist and their interactions. In this section, I'll discuss how measurements of the light element abundances can be used to constrain possible departures from the standard cosmological model.

8.6.1 Dark radiation

If any exotic radiation, decoupled from the Standard Model, was present during our universe's first few minutes, it would have impacted the expansion rate through the first Friedman equation, thereby altering the abundances of various nuclei that were produced during BBN. As we did in in Sec. 7.9, we will parameterize the energy density of dark radiation in terms of its contribution to $N_{\rm eff}$ and the energy density in photons:

$$\rho_{\rm DR} = \Delta N_{\rm eff} \times \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} \rho_\gamma, \tag{8.26}$$

where $\Delta N_{\rm eff} \equiv N_{\rm eff} - 3.046$.

The main way that the value of $\Delta N_{\rm eff}$ affects BBN is through its impact on the number of neutrons that are available to form nuclei. By increasing $\Delta N_{\rm eff}$, one increases the expansion

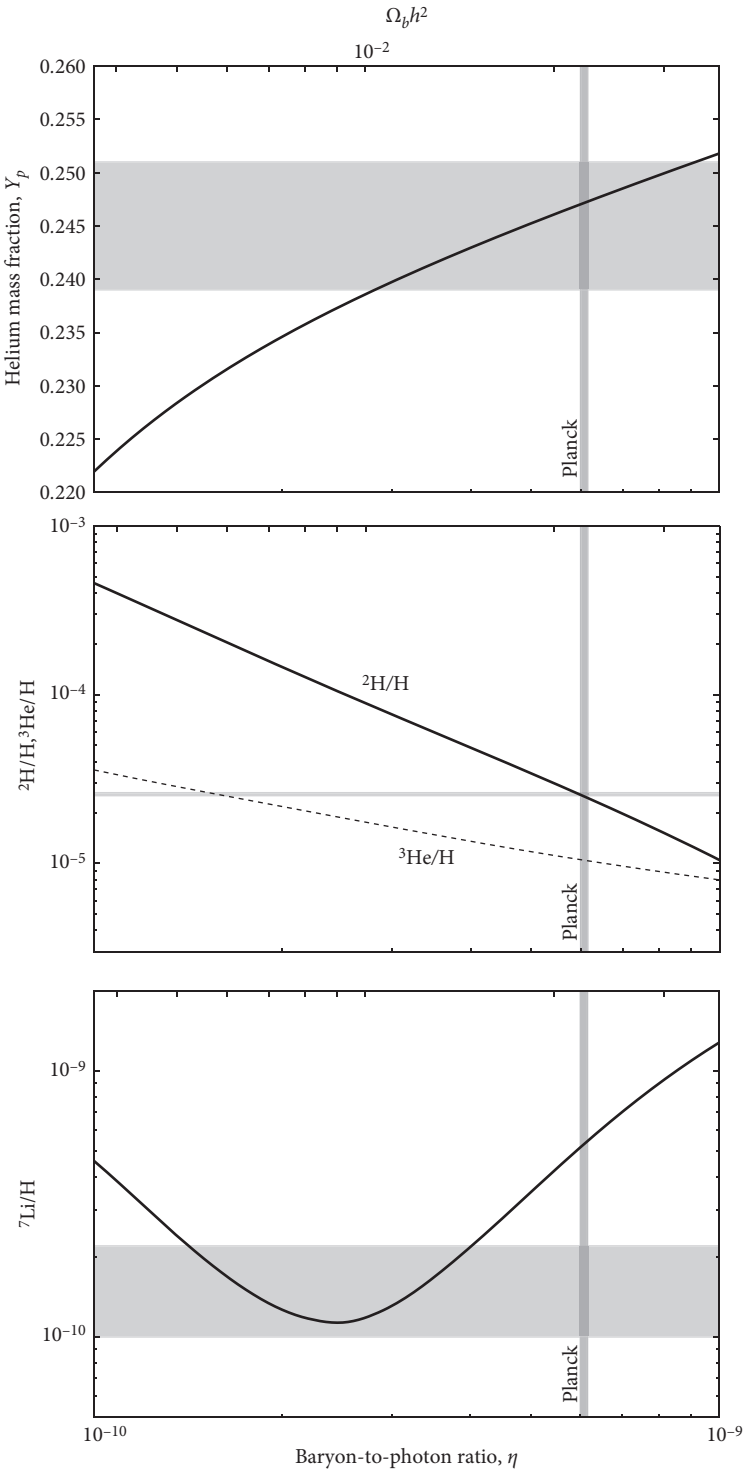


Figure 8.5 Measurements of the primordial element abundances compared to the values predicted by standard Big Bang nucleosynthesis (as calculated using the publicly available code PRIMAT), as a function of the baryon-to-photon ratio or the density of baryonic matter. All uncertainties shown are at the 95% confidence level. In the middle frame, the horizontal band represents the measured value of ^{2}H/H (the measured value of ^{3}He/H is not shown).

rate, causing the neutron-to-proton ratio to freeze out of equilibrium earlier, when the equilibrium neutron-to-proton ratio was higher. Although increasing the rate of expansion also works to suppress the effectiveness of subsequent fusion processes, this is a secondary effect. The ultimate impact of modifying the energy density in dark radiation is to change the helium mass fraction by $\Delta Y_p \approx 0.013 \times \Delta N_{\rm eff}$, and the relative deuterium abundance by $\Delta(^2{\rm H}/{\rm H})_p \approx (0.325 \times 10^{-5}) \times \Delta N_{\rm eff}$. After performing a fit to the measured values of the primordial helium and deuterium abundances in which both $N_{\rm eff}$ and $\Omega_b h^2$ are allowed to float independently, one finds that $2.3 < N_{\rm eff} < 3.4$ (at the 95% confidence level), consistent with the Standard Model prediction of 3.046.

Recall from Chapter 7 that the damping tail of the CMB's angular power spectrum is also sensitive to the value of $N_{\rm eff}$. Whereas the CMB provides us with a constraint on the energy density in radiation at around the time of recombination, measurements of the primordial element abundances are sensitive to the energy density that was present during our universe's first seconds and minutes. For a simple population of stable, relativistic particles, we expect their contribution to $N_{\rm eff}$ to be the same during both of these eras. In any number of other scenarios, however, the value of $N_{\rm eff}$ could be very different during BBN than it was during recombination. For this reason, these two ways of determining the value of $N_{\rm eff}$ are complementary, and independently probe different kinds of new physics.

8.6.2 Energy injection

In many scenarios involving physics beyond the Standard Model, energetic particles can be injected into the universe during or after the era of BBN, potentially impacting the primordial element abundances. For example, exotic particles could have been decaying during our universe's first minutes, creating high-energy photons and other particles that could be capable of disassociating nuclei.

An energetic photon can disassociate a nucleus in the early universe only if its energy exceeds the binding energy of its target and it is able to reach its target before being absorbed through the process of pair production. The rate for pair production can be written as

$$\Gamma = \int \sigma_{\gamma\gamma}(E_\gamma, \epsilon)\, c\, \frac{dn_\gamma}{d\epsilon}\, d\epsilon, \tag{8.27}$$

where E_γ is the energy of the nonthermal (i.e., the more energetic) photon, and $dn_\gamma/d\epsilon$ is the Bose-Einstein distribution of the thermal photons (see Eq. 4.4) in terms of their energy, ϵ. At energies well above the threshold for pair production, $E_\gamma \gg m_e^2/\epsilon$, the cross section for this process is roughly given by $\sigma_{\gamma\gamma} \sim 3\sigma_t m_e^2/2E_\gamma\epsilon$ (for a more detailed expression, see Sec. 5.3).

Consider, for example, a photon that has barely enough energy to disassociate a helium-4 nucleus, $E_\gamma = 28.3$ MeV. Such a photon can be absorbed by any photon with enough energy to exceed the threshold for pair production, $\epsilon \gtrsim m_e^2/E_\gamma \sim 9$ keV. Thus at temperatures greater than a few keV, the typical thermal photon encountered will have enough energy to absorb the higher-energy photon. Taking into account the high-energy tail of the Bose-Einstein distribution, the rate for pair production is high enough to severely limit the impact of photodisassociation down to temperatures of a fraction of a keV. At higher temperatures, nuclei can be broken up by other kinds of particles, such as energetic pions, but not efficiently by photons.

Once our universe has cooled enough to make photodisassociation potentially effective, any energetic photons that are injected into the bath can break up helium-4 nuclei. This process both reduces the abundance of helium-4, and generates additional deuterium. Roughly speaking, we can constrain or rule out such scenarios if the number of helium-disassociating

photons exceeds the uncertainty on the measured number of primordial deuterons. For example, if an exotic particle species with a lifetime of $\sim 10^8$ s produces one 30 MeV photon in each of its decays, and each of these photons produces one deuteron through the disassociation of a helium nucleus, then this would lead to a change in the deuteron abundance given by $\Delta(^2\mathrm{H}/\mathrm{H}) \sim (\rho_X/\rho_b)(m_p/m_X)$, where ρ_X and m_X are the density and mass of the decaying particle. If $\rho_X/\rho_b \gtrsim 3 \times 10^{-6} \times (m_X/100\,\mathrm{MeV})$, this would produce enough additional deuterium to increase its primordial abundance beyond the measured range. This can clearly lead to very strong constraints on particle species that are decaying or otherwise injecting energy into the universe during or shortly after the era of BBN.

Additional reading

Big Bang nucleosynthesis is discussed in some detail in Kolb and Turner's *The Early Universe*, as well as in Weinberg's *Cosmology*. There is also a useful and regularly updated review of this subject included in the Particle Data Group's *Review of Particle Physics* [24].

Problems

8.1 Why can massive stars produce elements like carbon, which were not formed in the early universe?

8.2 Make a back-of-the-envelope estimate for how much helium has been formed in stars over the course of our universe's history. How does this compare to the primordial helium abundance?

8.3 If the mass splitting between the neutron and the proton had been 12.93 MeV instead of 1.293 MeV, how would this have impacted the neutron-to-proton ratio over time? Estimate the temperature at which this ratio would have frozen out of equilibrium. How would this have impacted the final value of Y_p? Throughout this exercise, assume that the neutron's lifetime is equal to its standard value.

8.4 Iron-56 has a binding energy of 492 MeV, making it one of the most tightly bound nuclear species. If it had been included in the calculations described in Sec. 8.2, what would its equilibrium mass fraction have been as a function of temperature?

8.5 Using the result of the previous question, what was the rate of Hubble expansion at the temperature at which the equilibrium mass fraction of iron-56 became of order unity? Optimistically (and unrealistically) assume that most of the baryons were somehow in the form of chromium-48 and helium-4 at this time. What cross section would be required in order for the rate of this process to be equal to the Hubble rate? Is this realistic? Don't forget to consider the effects of coulomb repulsion.

8.6 What would happen if $\mathcal{O}(10\%)$ of the free nucleons in the sun were neutrons, as was the case in the early universe?

8.7 Write a code to numerically solve the Boltzmann equation for the synthesis of deuterium, ignoring all heavier nuclei. Plot the mass fractions of protons, neutrons, and deuterons as a function of temperature or time.

8.8 Consider a species of stable, relativistic, and decoupled fermions (with two internal degrees of freedom), with a temperature that is 10 times lower than that of the photons. What is its contribution to N_{eff}? Is this consistent with measurements of the primordial element abundances? Or with measurements of the CMB's angular power spectrum?

8.9 Assuming that the rates for proton-to-neutron scattering scale as $\Gamma \propto T^5$ during the time of neutron freeze out, calculate how the neutron-to-proton ratio (after freeze out) changes as a function of $N_{\rm eff}$. Then, assuming that nearly all of the neutrons end up in helium-4 nuclei, calculate how the resulting value of Y_p depends on that of $N_{\rm eff}$. Note: Because of the assumptions we have made in this problem, your answer should be similar, but not identical, to that cited in Sec. 8.6.1.

8.10 Consider an energetic photon that was injected into the universe when the temperature was 0.1 keV. Integrate over the thermal distribution of photons to determine the rate for pair production. What result do you get for $E_\gamma = 30$ MeV? What about $E_\gamma = 300$ MeV?

9

Dark Matter

So far, this book has focused on aspects of cosmology that are reasonably well understood. While we could certainly learn more by further studying the CMB or the light element abundances, no reasonable person would claim that we do not have a fairly robust understanding of how these relics of the Big Bang came to be formed. In this chapter, I depart from this trend, exploring instead one of the greatest mysteries in modern cosmology. Over the past several decades, astronomers and cosmologists have accumulated an overwhelming body of evidence that most of the matter in our universe consists of some exotic substance (or substances) that does not appreciably radiate, reflect, or absorb light. Despite having detected the presence of this matter in variety of different ways, we still do not know what this dark matter is, or how it was formed. This chapter summarizes the evidence for dark matter, and discusses a variety of ways in which it may have been produced in the early universe.

9.1 Dynamical evidence for dark matter

The earliest evidence for dark matter was based on observations of galaxy clusters. Such arguments go back to 1933, when Fritz Zwicky pointed out that the velocity dispersion of the galaxies in the Coma cluster is much higher than had been expected [47, 48]. If one assumes that the galaxies in a given cluster are in equilibrium and are isotropically distributed, then the velocity dispersion of those galaxies can be related to the total mass of the system. In particular, the virial theorem states that the time-averaged kinetic energy of such a system will be equal to half of the magnitude of its average gravitational potential energy. By applying this theorem to the case of the Coma cluster, it can be shown that the total mass of this system must be far greater than its mass in visible material, such as stars, gas, or dust.

In the decades that followed, similar behavior was observed among many other galaxy clusters. Doubts and questions remained, however, and no consensus was reached among astronomers regarding how to interpret these results. In particular, some argued that these clusters might not be in a state of equilibrium, but could instead contain galaxies that are not permanently bound to these systems.

This situation changed substantially in the 1970s, when a variety of new information came to light. In particular, astronomers including Vera Rubin and Kent Ford measured with

much greater precision the rates at which stars and gas orbit galaxies [49]. To the surprise of these and other astronomers, the rotation curves of the observed galaxies were found to be approximately "flat" out to the largest observed radii. In other words, the stars and gas were observed to be moving at roughly the same velocity, regardless of how far they were from their galaxy's center. These results implied that the observed galaxies are much more massive than expected, and that their mass is distributed in a way that extends well beyond the boundaries of the readily visible matter.

Around the same time, early numerical simulations began to show that rotationally supported galaxies consisting of a stellar disk should be unstable. In contrast to observations, such systems would form bars and rapidly evolve toward a more elliptical and pressure-supported configuration. In 1973, Jerry Ostriker and Jim Peebles proposed a solution to this problem, pointing out that a rotationally supported stellar disk would be stable if it were embedded within a massive spherical halo of dark matter [50].

In more recent decades, the observed dynamics of galaxies and galaxy clusters have continued to support the conclusion that our universe contains large quantities of dark matter. Furthermore, the existence of dark matter has been independently and robustly confirmed by a wide variety of cosmological observations. In particular, measurements of the CMB, the light element abundances, and the large-scale structure of our universe have each shown that dark matter—in particular, cold dark matter—makes up about 85% of our universe's total matter density.

9.2 MACHOs and gravitational microlensing

As early evidence for dark matter in galaxies and galaxy clusters accumulated, many astronomers and astrophysicists initially focused on the possibility that these systems could contain large abundances of relatively low-luminosity compact objects, such as brown dwarfs, red dwarfs, white dwarfs, planets, neutron stars, and black holes. This class of dark matter candidates would become known as MACHOs—massive astrophysical compact halo objects—in response to the competing classification of weakly interacting massive particles, WIMPs.

Through a combination of two main arguments, it was ultimately shown that (most types of) MACHOs cannot make up much of our universe's dark matter. First, observations of the CMB, the light element abundances, and large-scale structure have made it clear that the total density of matter is much larger than that of baryons, $\Omega_M \sim (6.3\text{–}6.5)\,\Omega_b$. With the exception of primordial black holes (which could have been formed prior to BBN, and thus do not necessarily contribute to the measured value of Ω_b), the limits on our universe's "baryon budget" exclude MACHOs as a major constituent of the dark matter. Second, astronomers have searched for MACHOs by conducting gravitational microlensing surveys, which are designed to detect the gravitational influence of compact objects on the light observed from luminous objects [51]. As a result of these surveys, it has become clear that the halos of the Milky Way and other nearby galaxies do not contain enough MACHOs to constitute the bulk of their dark matter, at least for MACHOs with masses in the range of $\sim 10^{-10} - 10\,M_\odot$.

Consider a nonluminous compact object that is perfectly positioned along the line of sight between an observer and a source of light, such as a distant star. This compact object will act as a gravitational lens, deflecting the light and causing the luminous object to appear as a ring of light, with an angular radius (known as the Einstein radius) equal to

the following:

$$\theta_E = \left(\frac{4GM}{c^2} \frac{(d_S - d_C)}{d_S d_C} \right)^{1/2}, \tag{9.1}$$

where M is the mass of the compact object, d_S is the distance to the star, and d_C is the distance to the compact object. For sources that are aligned with very massive objects, such as quasars lensed by galaxies or galaxy clusters, the effects of the gravitational lensing can be large enough to be resolved. Individual MACHOs, in contrast, generally induce a degree of deflection that is too small to be resolved by ground-based telescopes. A solar mass MACHO located at a distance of 1 kpc, for example, will lens the light from a 20 kpc distant star by only $\theta_E \approx 0.02$ arcseconds, more than an order of magnitude below the resolution of an idealized telescope located on the surface of the earth.

Even if the gravitational lensing from a MACHO cannot be resolved, this effect can lead to a detectable change in the apparent luminosity of a lensed star. As a MACHO passes in front of a star, the luminosity of that star is amplified by a factor given by

$$\mathcal{A} = \frac{(\theta/\theta_E)^2 + 2}{(\theta/\theta_E)[(\theta/\theta_E)^2 + 4]^{1/2}}, \tag{9.2}$$

where θ is the angle between the MACHO and the star. For an arbitrarily compact gravitational lens in perfect alignment, the amplification factor is infinite. In practice, amplification factors as large as several hundred have been observed.

Even if the entirety of the Milky Way's dark matter halo consisted of MACHOs, only approximately one out of two million stars would be significantly lensed ($\theta \lesssim \theta_E$) at any given time. This proportion is known as the microlensing optical depth. Each of these lensing events lasts only as long as the MACHO remains in approximate alignment along the line of sight. We can estimate the typical duration of such an event as follows:

$$\begin{aligned} t &\approx \frac{\theta_E \, d_C}{v_\perp} \\ &\approx \left(\frac{4GMd_C}{v_\perp^2 c^2} \right)^{1/2} \\ &\approx 140 \text{ days} \times \left(\frac{100 \text{ km/s}}{v_\perp} \right) \left(\frac{M}{M_\odot} \right)^{1/2} \left(\frac{d_C}{8 \text{ kpc}} \right)^{1/2}, \end{aligned} \tag{9.3}$$

where $v_\perp$ is the velocity of the MACHO perpendicular to the line of sight, and in the second line I have made the assumption that $d_S \gg d_C$. For practical reasons, it is easiest to search for microlensing events that last between roughly an hour and a few years, providing sensitivity to MACHOs with masses between $\sim 10^{-8}$ and $\sim 10^2 \, M_\odot$.

It would be impractical for astronomers to try to detect the presence of MACHOs by observing any one individual star. Even if the entirety of the dark matter were made up of MACHOs, one would have to wait for roughly $\sim (2 \times 10^6) \times t \sim 8 \times 10^5$ yrs $\times$ $(100 \text{ km s}^{-1}/v_\perp)(M/M_\odot)^{1/2}(d_C/8 \text{ kpc})^{1/2}$ before a MACHO would become aligned in such a way as to cause the observed star to be significantly lensed. Fortunately, it is possible for astronomers to observe large numbers of stars simultaneously, watching for the changes in brightness that could potentially result from gravitational microlensing.

Beginning in the 1990s, microlensing surveys carried out by the MACHO and EROS Collaborations reached the level of sensitivity needed to test the hypothesis that the dark

matter consists of compact objects. While these surveys were successful in detecting the presence of MACHOs in the halo of the Milky Way and elsewhere, the rate of these events was much lower than would have been expected if MACHOs had made up all of the dark matter. These results have thus been used to show that the abundance of MACHOs is well below the total density of dark mater, at least in the mass range of $M \sim 10^{-10} - 10\, M_{\odot}$ [52, 53, 54]. Whatever the dark matter is, it does not consist of compact objects within this range of masses.

9.3 What about modified gravity?

At least for the time being, the only uncontested evidence in support of dark matter is gravitational in nature. In other words, we infer the presence of dark matter in our universe uniquely through its gravitational effects. It is possible, for example, that the particles that make up this substance are entirely inert, and interact only gravitationally. From this perspective, one might also ask whether it is possible that dark matter does not exist at all, and instead we have misunderstood the nature of gravity.

Efforts to explain the phenomena ordinarily attributed to dark matter with modifications to gravity go back to 1982, when Mordehai Milgrom proposed the framework known as Modified Newtonian Dynamics, or MOND [55, 56, 57]. The basic idea behind MOND is that if one replaces Newton's second law, $F = ma$, with $F = ma^2/a_0$ in the very low-acceleration limit ($a \ll a_0 \sim 1.2 \times 10^{-10}\ \mathrm{m/s^2}$), then it would be possible to account for the observed rotation curves of galaxies, without the presence of any dark matter.

The original version of MOND did not respect the conservation of momentum, angular momentum, or energy, and it was not readily embedded within any theory consistent with general relativity. For these reasons and others, it was not viewed as a realistic theory, but rather as the approximate weak-field limit of some unknown and more physically viable theory. The most well-known and widely studied theory of MOND was proposed by Jacob Bekenstein in 2004, and goes by the name of tensor-vector-scalar gravity, or TeVeS [58]. Even this theory, however, has failed to account for the observed dynamics of galaxy clusters, and for the acoustic peaks in the angular power spectrum of the CMB. Observational problems for MOND were further elevated in 2006, when astronomers reported new measurements of a system known as the "bullet cluster," which consists of a pair of galaxy clusters that had recently undergone a collision [59]. As a result of this collision, the stars and galaxies within these clusters became spatially separated from the hot X-ray emitting gas that constitutes most of the baryonic mass in this system. When the X-ray maps are compared to the distribution of total mass as inferred through weak gravitational lensing, it becomes clear that the mass within the bullet cluster is displaced from the baryons. Another source of gravitational potential, such as that provided by dark matter, must instead dominate this system.

9.4 Requirements of a dark matter candidate

In order for a particle species to serve as a potentially viable candidate for dark matter, it must

- Be stable, or at least long-lived compared to the age of the universe,
- Be very feebly interacting with the particles of the Standard Model, and
- Be cold (i.e., nonrelativistic by the start of the matter-dominated era), in order be consistent with the observed large-scale structure of our universe (see Chapter 10).

None of the particles contained within the Standard Model satisfy all three of these requirements. The only Standard Model particles that are stable on cosmological timescales are the proton, electron, photon, and the three neutrinos. Of these, the proton, electron, and photon interact far too much to constitute viable dark matter candidates. Furthermore, the neutrinos produced in the Big Bang were still relativistic at the time of matter-radiation equality. In particular, the temperature of the neutrinos at this time was equal to $T_\nu^{\rm Eq} = T_{\rm CMB} \times (z_{\rm EQ}+1) \times (T_\nu/T_{\rm CMB}) \approx 0.000235\,{\rm eV} \times 3403 \times (4/11)^{1/3} \approx 0.57\,{\rm eV}$, which is larger than the mass of any of the three neutrino species.

In addition to the requirements listed above, we can place some rather general constraints on the mass of the individual constituents that make up the dark matter. First, a variety of dynamical considerations have ruled out dark matter candidates that are more massive than $M \sim 10$–$100\ M_\odot$ [60]. Furthermore, as discussed above, gravitational microlensing surveys have ruled out dark matter in the form of compact objects with masses in the range of $M \sim 10^{-10}$–$10\,M_\odot$. On the low-mass side, the particles that make up the dark matter cannot be so light as to prevent their de Broglie wavelength from fitting within the smallest halos that they collectively constitute. For the velocities typically found in dwarf galaxies, $v \sim 10\,{\rm km/s}$, a dark matter particle with a mass of 10^{-20} eV would have a de Broglie wavelength of $\lambda = h/p \sim 20\,{\rm pc}$, which is comparable to the half-light radii of the smallest observed galaxies. From the observed dynamics of these systems, one can rule out any dark matter candidates that are lighter than roughly $m \sim 10^{-20}$ eV [61, 62, 63, 64].

If the dark matter consists of fermions, its mass will be further restricted as a result of the Pauli exclusion principle, leading to what is known as the Tremaine-Gunn bound [65]. To see how this bound is obtained, consider a spheroidal galaxy of volume, V. For a uniform gas of fermions, the maximum number of these particles that can be present within this volume with $E < E_{\rm max}$ is bounded as follows:

$$N < \frac{V}{3\pi^2}\Big[2mE_{\rm max}\Big]^{3/2}, \tag{9.4}$$

where m is the mass of the fermionic dark matter particles. According to the virial theorem, the mass of this system is given by $M = \sigma^2 r/G$, where r and σ are the radius and velocity dispersion of the galaxy. Using $E_{\rm max} \approx m\sigma^2$ for the highest energy of the occupied states, this becomes

$$\frac{\sigma^2 r}{mG} \lesssim \frac{4r^3}{9\pi}\Big[2m^2\sigma^2\Big]^{3/2}, \tag{9.5}$$

which can be rearranged to obtain the following:

$$\begin{aligned} m &\gtrsim \left(\frac{9\pi}{8\sqrt{2}\,r^2\sigma G}\right)^{1/4} \\ &\gtrsim 0.6\,{\rm keV} \times \left(\frac{1\,{\rm kpc}}{r}\right)^{1/2}\left(\frac{10\,{\rm km/s}}{\sigma}\right)^{1/4}. \end{aligned} \tag{9.6}$$

When this result is applied to the smallest observed dwarf galaxies, one concludes that if the dark matter consists of a single fermionic particle species, its mass must be greater than roughly $\sim 1\,{\rm keV}$.

9.5 The abundance of a thermal relic

Any number of hypothetical particle species have been put forth as dark matter candidates, and many mechanisms have been proposed for the production of dark matter in the early universe. In this section, we're going to focus on the subset of these dark matter candidates that interacted enough in the early universe to have reached and maintained equilibrium with the Standard Model bath. Such particles are often referred to as thermal relics of the Big Bang.

Before moving on, let's pause and consider just how broad this subset of dark matter candidates is. As discussed in Sec. 4.4, all of the Standard Model particles interacted in the early universe at rates that were far greater than those required to keep them in a state of equilibrium. The first of these particles to decouple were the neutrinos, which froze out of equilibrium at around $T \sim \text{MeV}$, when the expansion rate became larger than their scattering rate. At very high temperatures, even very feeble interactions could potentially maintain a particle species in equilibrium with the Standard Model bath. Roughly speaking, a particle species will reach equilibrium with the Standard Model at a temperature T so long as its relevant scattering cross sections are larger than $\sigma \gtrsim 10^{-49}\,\text{cm}^2 \times (\text{TeV}/T)$ (see Sec. 4.5). This cross section is absolutely tiny! As a consequence, nearly any particle species that we could ever produce at a particle accelerator or otherwise detect in a laboratory would have easily reached equilibrium in the early universe.

From an initial condition of thermal and chemical equilibrium, we can calculate the evolution of the number density of a stable dark matter candidate, X, using the following Boltzmann equation (see Eq. 4.35):

$$\frac{dn_X}{dt} = -3Hn_X - \langle \sigma\, v \rangle \Big[n_X^2 - (n_X^{\text{Eq}})^2 \Big], \tag{9.7}$$

where $\langle \sigma\, v \rangle$ is the cross section for the annihilation of a pair of dark matter particles, thermally averaged as described in Sec. 4.7. As written, this equation describes the case in which the dark matter candidate is its own antiparticle (a Majorana particle), and the cross section is that for the process $XX \to SS'$, where S and S' denote Standard Model particles. Note that in this case the annihilation term includes two factors which cancel. The first is a factor of 1/2, which is required to avoid double counting the annihilation of particle A with particle B and vice versa, and the second is a factor of 2, which accounts for the fact that two dark matter particles are destroyed in each annihilation.

Alternatively, one could consider a Dirac dark matter candidate, which is distinct from its antiparticle.[1] In this case, we can use the following pair of coupled Boltzmann equations to describe the evolution of the number densities:

$$\frac{dn_X}{dt} = -3Hn_X - \langle \sigma\, v \rangle \Big[n_X n_{\bar{X}} - n_X^{\text{Eq}} n_{\bar{X}}^{\text{Eq}} \Big], \tag{9.8}$$

$$\frac{dn_{\bar{X}}}{dt} = -3Hn_{\bar{X}} - \langle \sigma\, v \rangle \Big[n_X n_{\bar{X}} - n_X^{\text{Eq}} n_{\bar{X}}^{\text{Eq}} \Big],$$

where this cross section describes the process $X\bar{X} \to SS'$. In the case that $n_X = n_{\bar{X}}$, we can combine these expressions to obtain an equation describing the evolution of the total

[1] In most situations, the labels Majorana and Dirac are reserved for fermions. In the context of annihilating dark matter, however, these words have come to refer to any particle species that is, or is not, its own antiparticle.

number density:

$$\begin{aligned}\frac{dn_{X+\bar{X}}}{dt} &= \frac{dn_X}{dt} + \frac{dn_{\bar{X}}}{dt} \\ &= -3Hn_X - 3Hn_{\bar{X}} - 2\langle\sigma\, v\rangle\left[n_X n_{\bar{X}} - n_X^{\rm Eq} n_{\bar{X}}^{\rm Eq}\right] \\ &= -3Hn_{X+\bar{X}} - \frac{1}{2}\,\langle\sigma\, v\rangle\left[n_{X+\bar{X}}^2 - (n_{X+\bar{X}}^{\rm Eq})^2\right],\end{aligned} \tag{9.9}$$

where $n_{X+\bar{X}} = n_X + n_{\bar{X}}$. Comparing this to Eq. 9.7, we find that we can calculate the evolution of a Dirac dark matter candidate's number density using the same Boltzmann equation as in the Majorana case, after making the substitution $\langle\sigma\, v\rangle \to \frac{1}{2}\langle\sigma\, v\rangle$.

The solutions to these Boltzmann equations correspond to the following series of events. At high temperatures, the annihilation and production rates are each much greater than that of Hubble expansion, $\langle\sigma\, v\rangle\, n_X^{\rm Eq} \gg H$, causing the abundance of the dark matter particles to remain very close to its equilibrium value. This continues until $\langle\sigma\, v\rangle\, n_X^{\rm Eq} \sim H$, at which point the dark matter particles freeze out of thermal equilibrium and are no longer efficiently created or destroyed. After freeze out, the number density of these particles simply evolves according to the increasing volume of the universe, $n_X \propto a^{-3}$.

9.5.1 Hot relics

The calculation of the thermal freeze out of a dark matter candidate depends on whether the particle species in question is relativistic or nonrelativistic at the time of its decoupling, and we'll discuss each of these cases one at a time. The final number density of dark matter particles is approximately given by their equilibrium abundance at the time of freeze out, diluted by three powers of the scale factor:

$$\begin{aligned}n_X(T) &\approx n_X^{\rm Eq}(T_f)\left(\frac{a}{a_f}\right)^{-3} \\ &= n_X^{\rm Eq}(T_f)\left(\frac{T}{T_f}\right)^3\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_f)}\right),\end{aligned} \tag{9.10}$$

where the subscript, f, denotes quantities evaluated at the time of freeze out and, in the second line, we have applied the conservation of entropy. If the dark matter particles are relativistic at the time of their freeze out, we can apply Eq. 4.6 to obtain

$$n_X(T) \approx \frac{\alpha\,\zeta(3)\,g_X\,T^3}{\pi^2}\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_f)}\right), \tag{9.11}$$

where g_X is the number of internal degrees of freedom associated with X, and $\alpha = 1$ or 3/4, depending on whether the dark matter is a boson or a fermion.

This result has wide reaching implications. In particular, it demonstrates that any hot relic will emerge from the Big Bang with a number density that is similar to the equilibrium number density of a light or massless particle, such as those that make up the CMB or the cosmic neutrino background (up to factors of g_X and $g_{\star,S}$). We can further use this

information to determine the energy density of these hot relics today:

$$\rho_X \approx \frac{\alpha\,\zeta(3)\,g_X\,T_{\rm CMB}^3}{\pi^2}\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)}\right) m_X \tag{9.12}$$
$$\approx 1.3\times10^{-6}\,{\rm GeV/cm^3}\times\left(\frac{\alpha}{0.75}\right)\left(\frac{g_X}{4}\right)\left(\frac{106.75}{g_{\star,S}(T_f)}\right)\left(\frac{m_X}{60\,{\rm eV}}\right).$$

Thus in order for a hot relic to freeze out with an abundance equal to the observed dark matter density, $\rho_{\rm DM} \equiv \Omega_{\rm DM}\,\rho_{\rm crit} \approx 1.3\times10^{-6}\,{\rm GeV/cm^3}$, it must be quite light, roughly in the $m_X \sim 1$–$100\,{\rm eV}$ range.

The canonical examples of a hot thermal relic are the three species of Standard Model neutrinos. As we found in Sec. 4.5, the neutrinos decoupled at a temperature of $T_f \sim {\rm MeV}$, at which time $g_{\star,S} \approx 10.75$. Combining this with $g_\nu = 2$ for each flavor, we arrive at

$$\rho_\nu \approx \sum_{i=1,2,3}\frac{3\zeta(3)\,T_{\rm CMB}^3}{2\pi^2}\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)}\right) m_{\nu_i} \tag{9.13}$$
$$\approx 3\times10^{-8}\,{\rm GeV/cm^3}\times\left(\frac{\sum_i m_{\nu_i}}{0.1\,{\rm eV}}\right).$$

This is well below the measured density of dark matter, even if the sum of the neutrino masses is equal to its current upper limit, $\sum_i m_{\nu_i} \approx 0.12\,{\rm eV}$.

9.5.2 Cold relics

Next, let's consider the thermal freeze out of a cold relic. To calculate the abundance of such a particle species, we start in the same way as we did in Eq. 9.10, but obtain a very different result:

$$n_X(T) \approx n_X^{\rm Eq}(T_f)\left(\frac{a}{a_f}\right)^{-3} \tag{9.14}$$
$$= n_X^{\rm Eq}(T_f)\left(\frac{T}{T_f}\right)^3\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_f)}\right)$$
$$= g_X\left(\frac{m_X T_f}{2\pi}\right)^{3/2} e^{-m_X/T_f}\left(\frac{T}{T_f}\right)^3\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_f)}\right).$$

From this, we can determine the current energy density of this cold relic:

$$\rho_X \approx g_X\left(\frac{m_X T_f}{2\pi}\right)^{3/2} e^{-m_X/T_f}\left(\frac{T_{\rm CMB}}{T_f}\right)^3\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)}\right) m_X \tag{9.15}$$
$$= g_X\,m_X\,T_{\rm CMB}^3\left(\frac{m_X}{2\pi\,T_f}\right)^{3/2}\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)}\right) e^{-m_X/T_f}.$$

Unlike in the case of a hot relic, the abundance of a cold relic depends very strongly on the temperature at which it freezes out, T_f. We can estimate this temperature by setting the

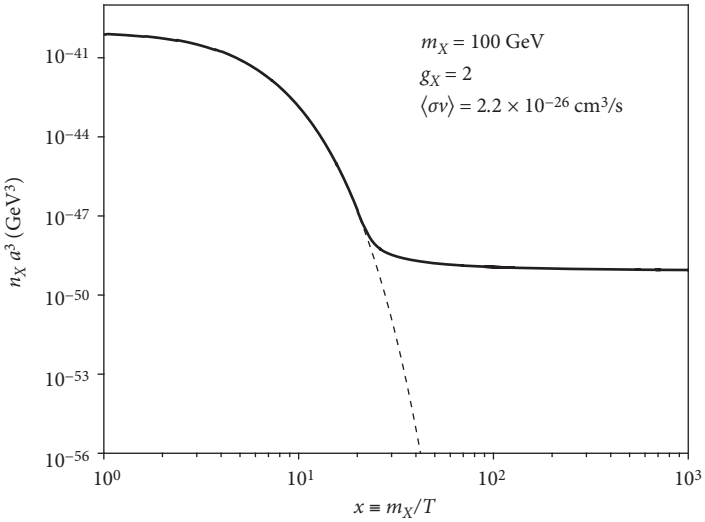


Figure 9.1 The comoving number density of a cold thermal relic, as a function of temperature. The case shown in that of a stable Majorana fermion with $g_X = 2$, $m_X = 100\,\text{GeV}$, and $\langle\sigma\, v\rangle = 2.2 \times 10^{-26}\,\text{cm}^3/\text{s}$. The dashed curve represents the equilibrium abundance.

annihilation rate per particle equal to the expansion rate:

$$1 = \frac{\langle\sigma\, v\rangle n_X^{\text{Eq}}}{H} = \frac{3\sqrt{10}\, g_X x_f^{3/2} T_f M_{\text{Pl}}\, e^{-x_f}\, \langle\sigma\, v\rangle}{8\pi^3 g_\star^{1/2}(T_f)}, \tag{9.16}$$

where in the last step I have introduced the quantity $x_f \equiv m_X/T_f$. Solving this expression for x_f, we find

$$\begin{aligned} x_f &= -\ln\left[\frac{8\pi^3 g_\star^{1/2}(T_f)}{3\sqrt{10}\, g_X x_f^{3/2} T_f M_{\text{Pl}}\, \langle\sigma\, v\rangle}\right] \\ &\approx 26 + \ln\left[\left(\frac{\langle\sigma\, v\rangle}{2.2\times 10^{-26}\,\text{cm}^3/\text{s}}\right)\left(\frac{80}{g_\star(T_f)}\right)^{1/2}\left(\frac{g_X}{2}\right)\left(\frac{x_f}{26}\right)^{3/2}\left(\frac{T_f}{10\,\text{GeV}}\right)\right]. \end{aligned} \tag{9.17}$$

Combining this expression for x_f with Eq 9.15, we obtain the following density:

$$\begin{aligned} \rho_X &\approx x_f\, T_{\text{CMB}}^3 \left(\frac{g_{\star,S}(T_{\text{CMB}})}{g_{\star,S}(T_f)}\right)\left(\frac{2\pi^{3/2} g_\star^{1/2}(T_f)}{3\sqrt{5}\, M_{\text{Pl}}\, \langle\sigma\, v\rangle}\right) \\ &\approx 1.3\times 10^{-6}\,\text{GeV/cm}^3 \times \left(\frac{2.2\times 10^{-26}\,\text{cm}^3/\text{s}}{\langle\sigma\, v\rangle}\right)\left(\frac{80}{g_\star}\right)^{1/2}\left(\frac{x_f}{26}\right). \end{aligned} \tag{9.18}$$

In these expressions, I have assumed that the dark matter candidate is Majorana, solving Eq. 9.7. If it is instead Dirac, the same results are obtained for a cross section that is twice as large. In the last line of Eq. 9.18, I assumed that $g_\star(T_f) = g_{\star,S}(T_f)$, which is valid so long as $T_f \gtrsim \text{MeV}$.

Despite the fact that I made the approximation of instantaneous freeze out in deriving this result, it is nonetheless quite accurate. The full numerical solution to the Boltzmann equation is shown in Figure 9.1, for the case of a Majorana fermion with two degrees of freedom, a mass of 100 GeV, and an annihilation cross section of $\langle \sigma v \rangle = 2.2 \times 10^{-26}$ cm^3/s. Until $T \sim m_X/20$, the abundance of these particles is maintained close to its equilibrium value, but then quickly transitions at freeze out to a state in which the comoving number density remains constant.

Notice from Eq. 9.18 that the final abundance of a cold thermal relic does not depend on the mass of the particle species, except logarithmically through its impact on x_f. Instead, this result depends most strongly on the value of the annihilation cross section. From this, you might be tempted to conclude that the dark matter could consist of a cold thermal relic with almost any mass. This is not the case, however. If the dark matter froze out after the start of BBN, the energy density of the dark matter and its annihilation products would have significantly altered the expansion rate and thus the primordial neutron-to-proton ratio, ruining the successful predictions for the light element abundances [66, 67]. On the other hand, the maximum possible value of a particle species' annihilation cross section is a function of its mass. From arguments based on partial wave unitarity, it can be shown that any dark matter candidate heavier than $\sim 10^2$ TeV will freeze out of equilibrium with an abundance that is greater than the observed dark matter density [68]. Taken together, these considerations require that, if the dark matter was in thermal equilibrium in the early universe, it must have a mass within the approximate range of ~ 3 MeV to $\sim 10^2$ TeV.

9.5.3 Velocity-dependent annihilations

So far in this discussion, we have taken the dark matter's annihilation cross section to be a simple numerical constant. In many dark matter models, however, σv is a function of the relative velocity of the annihilating particles, making the thermally averaged value, $\langle \sigma v \rangle$, dependent on the temperature.

Following the discussion around Eq. 4.31, the thermally averaged annihilation cross section for a nonrelativistic particle species is given by

$$\langle \sigma v \rangle \equiv \frac{\int \sigma v \, dn_1^{\rm Eq} dn_2^{\rm Eq}}{\int dn_1^{\rm Eq} dn_2^{\rm Eq}} \tag{9.19}$$

$$= \frac{\int \sigma v \, e^{-E_1/T} \, e^{-E_2/T} \, d^3\vec{p}_1 d^3\vec{p}_2}{\int e^{-E_1/T} \, e^{-E_2/T} \, d^3\vec{p}_1 d^3\vec{p}_2},$$

where the subscripts identify each of the two annihilating particles, and v is their Møller velocity (see Eq. 4.28). Even after a particle species has departed from chemical equilibrium, it will often remain in kinetic equilibrium, allowing us to write $f_i = f_i^{\rm Eq}$, up to an overall normalization factor. In the second line, I have further assumed that the particles are nonrelativistic and that they do not have any significant chemical potential, allowing us to write $f_i \approx e^{-E_i/T}$.

In this nonrelativistic case, we can rewrite these integrals in terms of the relative velocity of the two annihilating particles, $v = |\vec{p}_1 - \vec{p}_2|/m_X$, which simplifies the expression for the

thermally averaged cross section as follows:

$$\langle \sigma v \rangle \approx \frac{\int \sigma v \, v^2 \, e^{-xv^2/4} \, dv}{\int v^2 e^{-xv^2/4} \, dv} \tag{9.20}$$

$$= \frac{x^{3/2}}{2\pi^{1/2}} \int \sigma v \, v^2 \, e^{-xv^2/4} \, dv.$$

Note that to obtain this result, I used $(E_1 + E_2) = [(2m_X)^2 + (m_X v)^2]^{1/2} \approx 2m_X + m_X v^2/4$.

In many instances, it is convenient to write the annihilation cross section in terms of a Taylor series expansion in powers of v^2:

$$\sigma v = a + bv^2 + cv^4 + \cdots, \tag{9.21}$$

where a, b, and c are coefficients of this expansion. Processes arising from s-wave annihilation amplitudes contribute to all orders of this expansion, while p-wave amplitudes only contribute to the v^2 and higher order terms. For this reason, dark matter candidates that annihilate with a cross section that scales as $\sigma v \propto v^2$ are often referred to be being "p-wave suppressed."

For particles with an annihilation cross section given by $\sigma v = a + bv^2$, the integral in Eq. 9.20 yields the following result:

$$\langle \sigma v \rangle = \frac{x^{3/2}}{2\pi^{1/2}} \int (a + bv^2) \, v^2 e^{-xv^2/4} dv \tag{9.22}$$

$$= a + \frac{6b}{x}.$$

From this, we can use Eq. 9.17 determine the temperature at which freeze out occurs:

$$x_f = -\ln \left[\frac{8\pi^3 g_\star^{1/2}(T_f)}{3\sqrt{10}\, g_X x_f^{3/2} T_f M_{\rm Pl} \, (a + 6b/x_f)} \right]. \tag{9.23}$$

In order to determine the final abundance, it is useful to consider the probability that a given particle will annihilate after freeze out. This probability is given by

$$\int_{t_f}^{\infty} n_X \, \langle \sigma v \rangle \, dt \propto \int_{a_f}^{\infty} a^{-3} \langle \sigma v \rangle \, a \, da \propto \int_{x_f}^{\infty} \frac{\langle \sigma v \rangle}{x^2} dx, \tag{9.24}$$

where in the last steps I have made use of the facts that the comoving number density remains fixed after freeze out, $n_X \propto a^{-3}$, and that $dt \propto a \, da \propto x \, dx$ during radiation domination. If $\langle \sigma v \rangle$ is independent of temperature, the rightmost integral in the above expression is simply equal to $\langle \sigma v \rangle / x_f$. As expected, the relic abundance as given in Eq. 9.18 is inversely proportional to this combination of quantities. To generalize our calculation to the case of a temperature-dependent thermally averaged annihilation cross section, we simply replace

$\langle \sigma v \rangle / x_f$ in Eq. 9.18 with the following:

$$\frac{\langle \sigma v \rangle}{x_f} \longrightarrow \int_{x_f}^{\infty} \frac{\langle \sigma v \rangle}{x^2} dx \equiv J(x_f). \tag{9.25}$$

For the case of $\sigma v = a + b v^2$, this integral yields $J = (a + 3b/x_f)/x_f$. Combining this with Eq. 9.18, this leads to the following thermal relic abundance:

$$\begin{aligned} \rho_X &\approx T_{\rm CMB}^3 \left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)} \right) \left(\frac{2\pi^{3/2} g_\star^{1/2}(T_f)}{3\sqrt{5}\, M_{\rm Pl}\, J(x_f)} \right) \\ &\approx x_f\, T_{\rm CMB}^3 \left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)} \right) \left(\frac{2\pi^{3/2} g_\star^{1/2}(T_f)}{3\sqrt{5}\, M_{\rm Pl}\, (a + 3b/x_f)} \right). \end{aligned} \tag{9.26}$$

9.5.4 Annihilations near a resonance

Another class of commonly studied scenarios is that in which the dark matter particles annihilate through an s-channel resonance [69]. In such models, the annihilation cross section is enhanced when the total energy of the colliding particles in the center-of-momentum frame is approximately equal to the mass of the mediator. As an example, consider a scenario in which the dark matter's annihilation cross section takes the following form:

$$\sigma v = \frac{\alpha^2 s}{(m_{\rm med}^2 - s)^2 + m_{\rm med}^2 \Gamma_{\rm med}^2}, \tag{9.27}$$

where $m_{\rm med}$ and $\Gamma_{\rm med}$ are the mass and decay width of the particle that mediates the annihilation process, and α^2 normalizes the cross section. The Mandelstam variable, $s = (p_1 + p_2)^2 = E_{\rm CM}^2$, is related to the relative velocity of the two particles according to $s = 4m_X^2/(1 - v^2/4)$. To determine the thermally averaged cross section in this case, we insert the above expression into Eq. 9.20:

$$\langle \sigma v \rangle = \frac{x^{3/2}}{2\pi^{1/2}} \int \left(\frac{\alpha^2 s}{(m_{\rm med}^2 - s)^2 + m_{\rm med}^2 \Gamma_{\rm med}^2} \right) v^2 e^{-x v^2/4} dv. \tag{9.28}$$

Combining this with Eqs. 9.25 and 9.26, we can calculate the thermal relic abundance that results in this case.

To better understand how the presence of a resonance might impact the abundance of a dark matter candidate, let's consider a few illustrative cases. First, consider a dark matter candidate that annihilates far from the resonance, with $m_{\rm med} \gg 2m_X$. In this case, the annihilation cross section can be safely expanded in a Taylor series, $\sigma v \approx 4\alpha^2 m_X^2 (1 + v^2/4)/m_{\rm med}^4$, yielding an abundance as given in Eq. 9.26 for $a = 4\alpha^2 m_X^2/m_{\rm med}^4$ and $b = \alpha^2 m_X^2/m_{\rm med}^4$. Near resonance, however, the situation becomes more interesting. For the case of $m_{\rm med} = 2m_X$, the cross section grows rapidly as the velocity decreases, $\sigma v \propto v^{-4}$, approaching a maximum value of $\sigma v \simeq \alpha^2/\Gamma_{\rm med}^2$ in the low-velocity limit. After calculating the thermal average (and setting the value of α such that we obtain $\rho_X = \rho_{\rm DM}$), one finds that the low-temperature cross section can be significantly enhanced in this case, $\langle \sigma v \rangle_{T \to 0} \sim 2 \times 10^{-25}\, {\rm cm}^3/{\rm s} \times [(m_{\rm med}/\Gamma_{\rm med})/30]^2$. This result carries important implications for indirect dark matter searches (see Chapter 19). On the other hand, if a resonance is positioned such

that $m_{\rm med} \approx (2.05\text{–}2.10)\, m_X$, the dark matter will annihilate on resonance during freeze out, when $s \sim m_{\rm med}^2$. As a consequence, the annihilation cross section will become lower as the temperature of the dark matter decreases, suppressing the annihilation rate at late times.

9.5.5 Coannihilations

Another possibility is that, instead of being depleted through annihilations with other dark matter particles, the dark matter's relic abundance could be established through coannihilations with another particle species [69].

As an example, consider a stable dark matter candidate, X, that is accompanied by a heavier and unstable species, Y. The numbers of X and Y can be changed through the following interactions:

$$\begin{aligned} XX &\leftrightarrow S_i S_j, \\ YY &\leftrightarrow S_i S_j, \\ XY &\leftrightarrow S_i S_j, \\ XS_i &\leftrightarrow Y S_j, \\ Y &\leftrightarrow X S_i S_j, \end{aligned} \tag{9.29}$$

where S_i represent various species of Standard Model particles. Notice that in this example there are no interactions that can create or destroy a net number of X and Y particles in isolation. Instead, these particles are only destroyed or created in pairs.

The evolution of the number densities of X and Y are described by the following set of coupled Boltzmann equations:

$$\begin{aligned} \frac{dn_X}{dt} = &-3Hn_X - \sum_{S,S'} \langle \sigma_{XX\to SS'}\, v \rangle \left[n_X^2 - (n_X^{\rm Eq})^2 \right] \\ &- \sum_{S,S'} \langle \sigma_{XY\to SS'}\, v \rangle \left[n_X n_Y - n_X^{\rm Eq} n_Y^{\rm Eq} \right] \\ &- \sum_{S,S'} \langle \sigma_{XS\to YS'}\, v \rangle \left[n_X n_S \right] + \sum_{S,S'} \langle \sigma_{YS\to XS'}\, v \rangle \left[n_Y n_S \right] \\ &+ \sum_{S,S'} \Gamma_{Y\to XS_iS_j} \left[n_Y - n_Y^{\rm Eq} \frac{n_X}{n_X^{\rm Eq}} \right] \end{aligned} \tag{9.30}$$

and

$$\begin{aligned} \frac{dn_Y}{dt} = &-3Hn_Y - \sum_{S,S'} \langle \sigma_{YY\to SS'}\, v \rangle \left[n_Y^2 - (n_Y^{\rm Eq})^2 \right] \\ &- \sum_{S,S'} \langle \sigma_{XY\to SS'}\, v \rangle \left[n_X n_Y - n_X^{\rm Eq} n_Y^{\rm Eq} \right] \\ &- \sum_{S,S'} \langle \sigma_{YS\to XS'}\, v \rangle \left[n_Y n_S \right] + \sum_{S,S'} \langle \sigma_{XS\to YS'}\, v \rangle \left[n_X n_S \right] \\ &- \sum_{S,S'} \Gamma_{Y\to XS_iS_j} \left[n_Y - n_Y^{\rm Eq} \frac{n_X}{n_X^{\rm Eq}} \right]. \end{aligned} \tag{9.31}$$

Here, the sums are over the Standard Model species involved. The first interaction term in each case describes the processes of pair annihilation and pair production ($XX \leftrightarrow SS'$ or $YY \leftrightarrow SS'$), whereas the second such terms correspond to coannihilations between X and Y ($XY \leftrightarrow SS'$). The third and fourth interaction terms describe scattering process that convert an X into a Y or vice versa, and the final terms account for the decays and inverse decays of Y.

Before identifying the solution to this set of Boltzmann equations, let's pause to consider under what circumstances we might expect the species, Y, to play an important role in determining the final abundance of our dark matter candidate, X. The relative abundance of these states at the time of dark matter freeze out can be estimated as follows:

$$\frac{n_Y}{n_X} \sim \frac{n_Y^{\rm Eq}}{n_X^{\rm Eq}} \approx \frac{g_Y}{g_X} \frac{e^{-m_Y/T_f}}{e^{-m_X/T_f}} = \frac{g_Y}{g_X} e^{-\Delta x_f}, \tag{9.32}$$

where $\Delta \equiv (m_Y - m_X)/m_X$ is the fractional mass splitting between the two states. For large mass splittings, $\Delta \gg 1/x_f \equiv T_f/m_X \sim 1/25$, there will be far fewer Y particles than dark matter particles at the time of freeze out, preventing the Y population from playing any significant role in this process. If the mass splitting is smaller, however, coannihilations could potentially be the primary way in which the X population is depleted in the early universe, in particular if $\langle \sigma_{XY \to SS'} v \rangle \gg \langle \sigma_{XX \to SS'} v \rangle$.

To determine the dark matter's thermal relic abundance in this class of scenarios, one could use a brute force approach and solve the set of coupled Boltzmann equations described above numerically. Alternatively, this problem can be handled analytically by considering the combined number density of X and Y particles. This quantity is described by the following Boltzmann equation:

$$\begin{aligned} \frac{dn_{X+Y}}{dt} = -3Hn_{X+Y} &- \sum_{S,S'} \langle \sigma_{XX \to SS'} v \rangle \left[n_X^2 - (n_X^{\rm Eq})^2 \right] \\ &- \sum_{S,S'} \langle \sigma_{YY \to SS'} v \rangle \left[n_Y^2 - (n_Y^{\rm Eq})^2 \right] \\ &- \sum_{S,S'} \langle \sigma_{XY \to SS'} v \rangle \left[n_X n_Y - n_X^{\rm Eq} n_Y^{\rm Eq} \right]. \end{aligned} \tag{9.33}$$

Notice that I have not included in this expression any interactions of the type $XS_i \leftrightarrow YS_j$ or $Y \leftrightarrow XS_iS_j$, as these do not change the value of n_{X+Y}. Also bear in mind that it is n_{X+Y}, as opposed to n_X, that will determine the final dark matter abundance, since any Ys that survive freeze out will ultimately decay to produce an X.

Efficient scattering of the form $XS_i \leftrightarrow YS_j$ will ensure that the quantities n_X/n_{X+Y} and n_Y/n_{X+Y} remain near their equilibrium values throughout the process of freeze out. Under this approximation, we can rewrite Eq. 9.33 in the form:

$$\begin{aligned} \frac{dn_{X+Y}}{dt} = -3Hn_{X+Y} &- \sum_{S,S'} \langle \sigma_{XX \to SS'} v \rangle \left[n_{X+Y}^2 - (n_{X+Y}^{\rm Eq})^2 \right] \left(\frac{n_X^{\rm Eq}}{n_{X+Y}^{\rm Eq}} \right)^2 \\ &- \sum_{S,S'} \langle \sigma_{YY \to SS'} v \rangle \left[n_{X+Y}^2 - (n_{X+Y}^{\rm Eq})^2 \right] \left(\frac{n_Y^{\rm Eq}}{n_{X+Y}^{\rm Eq}} \right)^2 \end{aligned}$$

$$- \sum_{S,S'} \langle \sigma_{XY \to SS'} v \rangle \left[n_{X+Y}^2 - (n_{X+Y}^{\rm Eq})^2 \right] \left(\frac{n_X^{\rm Eq} n_Y^{\rm Eq}}{(n_{X+Y}^{\rm Eq})^2} \right), \tag{9.34}$$

which after some rearrangement can be written as follows:

$$\frac{dn_{X+Y}}{dt} = -3Hn_{X+Y} - \sum_{S,S'} \left[\langle \sigma_{XX \to SS'} v \rangle \left(\frac{n_X^{\rm Eq}}{n_{X+Y}^{\rm Eq}} \right)^2 + \langle \sigma_{YY \to SS'} v \rangle \left(\frac{n_Y^{\rm Eq}}{n_{X+Y}^{\rm Eq}} \right)^2 \right. \tag{9.35}$$
$$\left. + \langle \sigma_{XY \to SS'} v \rangle \left(\frac{n_X^{\rm Eq} n_Y^{\rm Eq}}{(n_{X+Y}^{\rm Eq})^2} \right) \right] \left[n_{X+Y}^2 - (n_{X+Y}^{\rm Eq})^2 \right].$$

We now have a Boltzmann equation that is in a form we are familiar with. In particular, we can think of the quantity in the large square brackets as the effective cross section for the combined effects of annihilation and coannihilation. This effective cross section can be written as follows:

$$\langle \sigma_{\rm eff} v \rangle = \sum_{S,S'} \left[\langle \sigma_{XX \to SS'} v \rangle \, \frac{g_X^2}{g_{\rm eff}^2} + \langle \sigma_{YY \to SS'} v \rangle \, \frac{g_Y^2}{g_{\rm eff}^2} \, (1+\Delta)^3 \, e^{-2x\Delta} \right. \tag{9.36}$$
$$\left. + 2 \langle \sigma_{XY \to SS'} v \rangle \, \frac{g_X g_Y}{g_{\rm eff}^2} \, (1+\Delta)^{3/2} \, e^{-x\Delta} \right],$$

where

$$g_{\rm eff} \equiv g_X + g_Y (1+\Delta)^{3/2} \, e^{-x\Delta}. \tag{9.37}$$

This calculation can be straightforwardly generalized to describe scenarios that include an arbitrary number of annihilating and coannihilating species. In this case, we use the following Boltzmann equation:

$$\frac{dn}{dt} = -3Hn - \langle \sigma_{\rm eff} v \rangle \left[n^2 - (n^{\rm Eq})^2 \right], \tag{9.38}$$

where $n = n_X + n_Y + \cdots$ is the total abundance of all annihilating/coannihilating particles, and the effective cross section is given by

$$\langle \sigma_{\rm eff} v \rangle = \sum_{i,j} \langle \sigma_{ij} v \rangle \frac{g_i g_j}{g_{\rm eff}^2} \, (1+\Delta_i)^{3/2} \, (1+\Delta_j)^{3/2} \, e^{-x(\Delta_i + \Delta_j)}, \tag{9.39}$$

where $\langle \sigma_{ij} v \rangle$ is the cross section for the process $X_i X_j \to SS'$ summed over all Standard Model final states, g_i is the number of internal degrees of freedom of state i, and $\Delta_i \equiv (m_i - m_1)/m_1$ is the mass splitting between state i and the lightest annihilating/coannihilating state. The effective number of degrees of freedom in this expression is given by

$$g_{\rm eff} \equiv \sum_i g_i (1+\Delta_i)^{3/2} e^{-x\Delta_i}. \tag{9.40}$$

Employing the effective cross section given in Eq. 9.39, we can use Eq. 9.18 to determine the thermal relic abundance of the lightest annihilating/coannihilating species.

Before moving on, let's use this framework to consider a couple of illustrative examples. In the case of two states with the same number of degrees of freedom ($g_X = g_Y$) and with a very small mass splitting ($\Delta \ll 1/x_f$), the effective cross section reduces to $\langle \sigma_{\rm eff}\, v\rangle \approx 0.25\,\langle \sigma_{XX} v\rangle + 0.25\,\langle \sigma_{YY} v\rangle + 0.5\,\langle \sigma_{XY} v\rangle$. In the limit that $\langle \sigma_{XY} v\rangle \gg \langle \sigma_{XX} v\rangle, \langle \sigma_{YY} v\rangle$, coannihilations will play the dominant role in depleting the combined abundance of $X + Y$. In the opposite case, $\langle \sigma_{XX} v\rangle, \langle \sigma_{YY} v\rangle \gg \langle \sigma_{XY} v\rangle$, the X and Y populations will each freeze out independently, after which each Y decays to create an X.

9.6 Asymmetric dark matter

Let's begin this section by taking the approach described earlier in this chapter to estimate the thermal relic abundance of baryons in our universe. From the measured cross section for proton-antiproton annihilation, $\sigma_{p\bar{p}} v \sim \mathcal{O}(10^{-15}\ {\rm cm}^3/{\rm s})$, we can use Eqs. 9.17 and 9.18 to show that these particles should have undergone thermal freeze out at a temperature around $T_f \sim 20\,{\rm MeV}$, yielding an abundance of roughly $\rho_b \sim 10^{-16}\,{\rm GeV/cm}^3$. This result corresponds to $\Omega_b \sim 10^{-11}$, which is many orders of magnitude smaller than the measured value of $\Omega_b \approx 0.049$.

This result has forced cosmologists to conclude that there must have been an asymmetry between the abundances of matter and antimatter in the early universe (see Chapter 12). In other words, for reasons that are not yet understood, the early universe contained slightly more baryons than antibaryons (and more quarks than antiquarks prior to the QCD phase transition). As a consequence of this asymmetry, these particles did not undergo the process of thermal freeze out in the way that was described in the previous section. Instead of freezing out of equilibrium at $T \sim 20\,{\rm MeV}$, the comoving abundance of these particles became fixed at an earlier time, when the annihilations had destroyed all of the antibaryons, leaving behind a residual population of relic baryons.

It is possible that there may have also been a primordial asymmetry between the number of dark matter particles, X, and their antiparticles, $\bar{X}$ (see Refs. [70, 71, 72] and references therein). In the previous section, I explicitly assumed that $n_X = n_{\bar{X}}$. Generalizing the Boltzmann equation to the case in which $n_X \neq n_{\bar{X}}$, we have the following:

$$\frac{dn_X}{dt} = -3Hn_X - \langle \sigma\, v\rangle (n_X n_{\bar{X}} - n_X^{\rm Eq} n_{\bar{X}}^{\rm Eq}), \tag{9.41}$$
$$\frac{dn_{\bar{X}}}{dt} = -3Hn_{\bar{X}} - \langle \sigma\, v\rangle (n_X n_{\bar{X}} - n_X^{\rm Eq} n_{\bar{X}}^{\rm Eq}).$$

To account for the asymmetry between the X and $\bar{X}$ abundances, we introduce a chemical potential, $\mu_X = -\mu_{\bar{X}}$. As we have seen before, the abundances of these particles freeze out of equilibrium when the rate of Hubble expansion exceeds their rate of annihilation. But in this case, the chemical potential suppresses the value of $n_{\bar{X}}^{\rm Eq}$, and thus the annihilation rate. Setting the annihilation rate equal to $Hn_X^{\rm Eq}$, we have the following condition for freeze out:

$$Hn_X^{\rm Eq} \approx \langle \sigma\, v\rangle n_X^{\rm Eq} n_{\bar{X}}^{\rm Eq} \tag{9.42}$$
$$\longrightarrow \left(\frac{8\pi^3 g_\star(T_f)\, T_f^4}{90 M_{\rm Pl}^2}\right)^{1/2} \approx \langle \sigma\, v\rangle\, g_X \left(\frac{m_X T_f}{2\pi}\right)^{3/2} e^{-(m_X - \mu_{\bar{X}})/T_f},$$

which yields the following freeze-out temperature:

$$x_f \equiv \frac{m_X}{T_f} \approx -\ln\left[\frac{8\pi^3 g_\star^{1/2}(T_f)}{3\sqrt{10}\, g_X\, x_f^{3/2}\, T_f\, M_{\rm Pl}\langle\sigma v\rangle}\right] + \frac{\mu_{\bar{X}}}{T_f}. \tag{9.43}$$

$$\approx 25 + \ln\left[\left(\frac{\langle\sigma v\rangle}{4.4\times10^{-26}\,{\rm cm}^3/{\rm s}}\right)\left(\frac{80}{g_\star(T_f)}\right)^{1/2}\left(\frac{g_X}{2}\right)\left(\frac{x_f}{26}\right)^{3/2}\left(\frac{T_f}{10\,{\rm GeV}}\right)\right] - \frac{\mu_X}{T_f}.$$

From this expression, we see that the effect of the dark matter's chemical potential is to reduce the value of x_f, and thus to cause freeze out to occur earlier. In the presence of a significant chemical potential, the final dark matter abundance is given by (see Eqs. 9.15 and 9.18):

$$\rho_X \approx g_X\, m_X\, T_{\rm CMB}^3 \left(\frac{m_X}{2\pi\, T_f}\right)^{3/2}\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)}\right) e^{-m_X/T_f} \tag{9.44}$$

$$\approx x_f\, T_{\rm CMB}^3 \left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_f)}\right)\left(\frac{2\pi^{3/2} g_\star^{1/2}(T_f)}{3\sqrt{5}\, M_{\rm Pl}\,\langle\sigma v\rangle}\right) e^{\mu_X/T_f}$$

$$\approx 1.3\times10^{-6}\,{\rm GeV/cm}^3 \times \left(\frac{2.2\times10^{-26}\,{\rm cm}^3/{\rm s}}{\langle\sigma v\rangle}\right)\left(\frac{80}{g_\star(T_f)}\right)^{1/2}\left(\frac{x_f}{26}\right) e^{\mu_X/T_f}.$$

If $\mu_X \ll T_f$, the particles freeze out normally, and the asymmetry does not significantly impact their resulting abundance. In this case, the dark matter consists of approximately equal quantities of X and $\bar{X}$ particles, with $\rho_{\rm DM} = \rho_X + \rho_{\bar{X}} \approx 2\rho_X$. In order for this combined abundance to equal the measured dark matter density, the dark matter's annihilation cross section must be $\langle\sigma v\rangle \approx 4.4\times10^{-26}\,{\rm cm}^3/{\rm s}$. In contrast, if $\mu_X \gg T_f$, $n_{\bar{X}}$ will become much smaller than n_X prior to freeze out, suppressing the annihilation rate and making freeze out occur earlier than it would have in the absence of a chemical potential. Consider, for example, a dark matter candidate with an annihilation cross section that is 100 times larger than that associated with a typical (Dirac) relic, $\langle\sigma v\rangle \approx 4.4\times10^{-24}\,{\rm cm}^3/{\rm s}$. If the dark matter has no chemical potential, it will freeze out with an abundance of $\Omega_{X+\bar{X}} \sim 10^{-2}\,\Omega_{\rm DM}$. If the dark matter instead has a chemical potential of $\mu_X \approx 0.22\, m_X$, the asymmetry will cause freeze out to occur at a temperature of $T_f \approx 0.041 m_X$, despite the large annihilation cross section. In this case, the factor of $e^{\mu_X/T_f} \approx 200$ in Eq. 9.44 offsets the factor resulting from the large annihilation cross section, leading us to obtain a relic abundance that is consistent with the measured dark matter density.

One motivation for this class of scenarios is the possibility that there could exist a dark matter asymmetry that is somehow related to the baryon asymmetry. In particular, one could imagine models in which the dark matter carries a negative quantity of baryon number (but not QCD color), with a total magnitude that is exactly equal to that carried by quarks, making our overall universe neutral with respect to baryon number. Such scenarios are attractive in that they could potentially explain why $\Omega_{\rm DM}$ is of the same order of magnitude as Ω_b, which is simply a coincidence in most dark matter models. In order for the baryon number carried by the dark matter to cancel that of the quarks found in our universe, $|q_b|\, n_X = n_b$ must be satisfied, where q_b is the (negative) value of the baryon number assigned to the dark matter candidate. This, in turn, allows us to determine the

mass of the dark matter candidate, as a function of q_b:

$$m_X = \frac{\Omega_{\rm DM}}{\Omega_b}\, m_{p,n}\, |q_b| \approx 5.1\ {\rm GeV} \times |q_b|. \tag{9.45}$$

While many asymmetric dark matter models have been proposed in which this relationship is not strictly realized [70, 71, 72], the number densities of dark matter particles and baryons are generally expected to be comparable in this class of scenarios, favoring dark matter candidates with masses in the range of roughly $m_X \sim 1$–15 GeV, depending on details of the model under consideration.

9.7 Nonthermal production mechanisms

In Secs. 9.5 and 9.6, we assumed that the dark matter was in chemical and kinetic equilibrium with the Standard Model bath at some point in the early universe. If the dark matter is extremely feebly interacting, however, it is possible that its relevant scattering rates may have never exceeded the rate of Hubble expansion, thus never bringing these particles into equilibrium.

There are many ways that dark matter could have been produced in the early universe without reaching equilibrium. Some of these possibilities we will explore in later chapters, such as the misalignment mechanism for the production of axions (see Sec. 17.3). In this section, we will explore two other possible mechanisms for the nonthermal production of dark matter: freeze in and out-of-equilibrium decays.

9.7.1 Freeze in

Even if the couplings of the dark matter to the Standard Model are too feeble to bring these particles into chemical equilibrium, collisions of particles in the thermal bath could have generated a significant abundance of dark matter in the early universe. This class of scenarios is generally referred to as dark matter "freeze in" [73].

Let's begin by considering a particle species that can be produced in pairs in the collisions of Standard Model particles, $SS' \to XX$, but with a cross section that is so small that the rate of Hubble expansion is always much larger than the scattering rate, $H \gg \langle \sigma\, v \rangle n_X^{\rm Eq}$, even at very high temperatures. Let's further assume that the universe did not contain any significant abundance of X particles at very early times. We can determine how the number density of this species evolves over time by solving the appropriate Boltzmann equation (Eq. 9.7) in the $n_X \ll n_X^{\rm Eq}$ limit:

$$\frac{dn_X}{dt} = -3Hn_X + \langle \sigma\, v \rangle (n_X^{\rm Eq})^2, \tag{9.46}$$

which we can rewrite as follows:

$$\frac{d(a^3 n_X)}{dt} = a^3 \langle \sigma\, v \rangle (n_X^{\rm Eq})^2. \tag{9.47}$$

It is useful in this context to consider the change in the comoving number density per Hubble time, $t_H \equiv H^{-1}$:

$$\begin{aligned}\frac{d(a^3 n_X)}{dt_H} &\propto \frac{1}{H}\frac{d(a^3 n_X)}{dt} \\ &\propto \frac{a^3 \langle \sigma\, v\rangle (n_X^{\rm Eq})^2}{H} \\ &\propto \langle \sigma\, v\rangle\, T.\end{aligned} \tag{9.48}$$

From this calculation, we see that if the cross section is independent of temperature, the largest contributions to the final X abundance will be generated when the temperature is highest. For this reason, the resulting dark matter abundance in this scenario is highly sensitive to the initial conditions, such as the temperature and state of the universe immediately after inflation.

Scenarios for the production of dark matter that are sensitive to high-temperature initial conditions are said to be "UV-dominated." One could also, however, consider cases in which the dark matter is produced more efficiently at lower temperatures, and thus at later times. Consider, for example, the production of a scalar dark matter candidate through the s-channel exchange of a spin-1 mediator, $SS' \to Z' \to XX$. In the limit of a light mediator mass, $m_{Z'} \ll T, m_X$, the cross section for this process scales as $\langle \sigma\, v\rangle \propto T^{-2}$. In this case, the comoving number density evolves as $d(a^3 n_X)/dt_H \propto (\sigma\, v)_0\, T^{-1}$, where $\langle \sigma\, v\rangle = (\sigma\, v)_0\, T^{-2}$. This scaling continues until the temperature drops below the mass of the dark matter candidate, and therefore the Xs are most efficiently produced at relatively modest temperatures, $T \sim m_X$, making this class of scenarios "IR-dominated."

We can solve the Boltzmann equation in this case to determine the evolution of n_X. In the limit of $T \gg m_X$, this is given by:

$$\begin{aligned}\frac{d(a^3 n_X)}{dt} &= a^3 \langle \sigma\, v\rangle (n_X^{\rm Eq})^2 \\ &= \left(\frac{T_{\rm CMB}}{T}\right)^3 \left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T)}\right) \frac{(\sigma\, v)_0}{T^2} \left(\frac{g_X \zeta(3) T^3}{\pi^2}\right)^2 \\ &= \frac{g_X^2\, \zeta(3)^2\, T_{\rm CMB}^3\, g_{\star,S}(T_{\rm CMB})\, (\sigma\, v)_0\, T}{\pi^4\, g_{\star,S}(T)},\end{aligned} \tag{9.49}$$

where in the second line I have used the conservation of entropy to express the scale factor in terms of temperature. Using this again to write $dt \approx -dT/TH$ (up to variations in $g_{\star,S}$; see Problem 9.10), we can solve the above expression for $a^3 n_X$:

$$\begin{aligned}a^3 n_X &\sim \frac{g_X^2\, \zeta(3)^2\, T_{\rm CMB}^3\, g_{\star,S}(T_{\rm CMB})\, (\sigma\, v)_0}{\pi^4} \int_{T_{\rm min}}^{\infty} \frac{dT}{H g_{\star,S}(T)} \\ &= \frac{3\sqrt{5}\, g_X^2\, \zeta(3)^2\, T_{\rm CMB}^3\, g_{\star,S}(T_{\rm CMB})\, (\sigma\, v)_0\, M_{\rm Pl}}{2\,\pi^{11/2}} \int_{T_{\rm min}}^{\infty} \frac{dT}{g_\star^{1/2}(T)\, g_{\star,S}(T)\, T^2}\end{aligned} \tag{9.50}$$

$$\sim \frac{3\sqrt{5}\, g_X^2 \zeta(3)^2\, T_{\rm CMB}^3\, g_{\star,S}(T_{\rm CMB})\, (\sigma v)_0\, M_{\rm Pl}}{2\,\pi^{11/2}\, g_\star^{1/2}(T_{\rm min})\, g_{\star,S}(T_{\rm min})\, T_{\rm min}},$$

where I have integrated up to the point at which the dark matter becomes non-relativistic, $T_{\rm min} \sim m_X$. We can use this result to estimate the resulting dark matter abundance in the universe today:

$$\begin{aligned}\rho_X &= n_X\, m_X \qquad (9.51)\\ &\sim \frac{3\sqrt{5} g_X^2 \zeta(3)^2\, T_{\rm CMB}^3\, g_{\star,S}(T_{\rm CMB})\, (\sigma v)_0\, M_{\rm Pl}}{2\,\pi^{11/2}\, g_\star^{1/2}(T_{\rm min})\, g_{\star,S}(T_{\rm min})}\, \frac{m_X}{T_{\rm min}}\\ &\sim 1.3\times 10^{-6}\,{\rm GeV/cm^3}\,\times \left(\frac{(\sigma v)_0}{2\times 10^{-24}}\right)\left(\frac{g_X}{1}\right)^2\left(\frac{100}{g_{\star,S}(T_{\rm min})}\right)\left(\frac{100}{g_\star(T_{\rm min})}\right)^{1/2}\left(\frac{m_X}{T_{\rm min}}\right).\end{aligned}$$

Instead of taking the relativistic limit for $n_X^{\rm Eq}$, we can get a more accurate result by calculating $n_X^{\rm Eq}$ numerically and using this in the integration of Eq. 9.49:

$$\begin{aligned}\rho_X &= n_X\, m_X \qquad (9.52)\\ &\approx \frac{3\sqrt{5}\, T_{\rm CMB}^3\, g_{\star,S}(T_{\rm CMB})\, (\sigma v)_0\, M_{\rm Pl}}{2\pi^{3/2}} \int_0^\infty \frac{(n_X^{\rm Eq})^2\, dT}{g_\star^{1/2}(T)\, g_{\star,S}(T)\, T^8}\\ &\approx 1.3\times 10^{-6}\,{\rm GeV/cm^3}\,\times \left(\frac{(\sigma v)_0}{1.4\times 10^{-25}}\right)\left(\frac{g_X}{1}\right)^2\left(\frac{100}{g_{\star,S}(T_{\rm min})}\right)\left(\frac{100}{g_\star(T_{\rm min})}\right)^{1/2}.\end{aligned}$$

From this, we learn that the dark matter could be produced through freeze in, even if its couplings to the Standard Model are really quite small, $\langle\sigma v\rangle \sim (1.4\times 10^{-25})/T^2 \approx 1.6\times 10^{-42}\,{\rm cm^3/s}\times({\rm GeV}/T)^2$.

9.7.2 Out-of-equilibrium decay

As another example, we will consider a heavy and relatively long-lived particle species, Y, which is present with some abundance in the very early universe. If Y is very feebly interacting, it will either start out decoupled from the Standard Model bath, or otherwise decouple from it at some point as the universe expands and cools. After decoupling, the energy density of these particles evolves as follows:

$$\rho_Y = \begin{cases} \rho_{Y,i}\left(\dfrac{a}{a_i}\right)^{-4} = \rho_{Y,i}\left(\dfrac{T}{T_i}\right)^4\left(\dfrac{g_{\star,S}(T)}{g_{\star,S}(T_i)}\right)^{4/3}, & \rho_Y \gg n_Y m_Y,\\ n_{Y,i}\, m_Y\left(\dfrac{a}{a_i}\right)^{-3} = n_{Y,i}\, m_Y\left(\dfrac{T}{T_i}\right)^3\left(\dfrac{g_{\star,S}(T)}{g_{\star,S}(T_i)}\right), & \rho_Y \approx n_Y m_Y, \end{cases} \qquad (9.53)$$

where quantities with the subscript i represent the initial values, and the upper and lower expressions apply to periods in which the particles are relativistic and nonrelativistic, respectively. We can use this to find how ρ_Y changes relative to the Standard Model energy

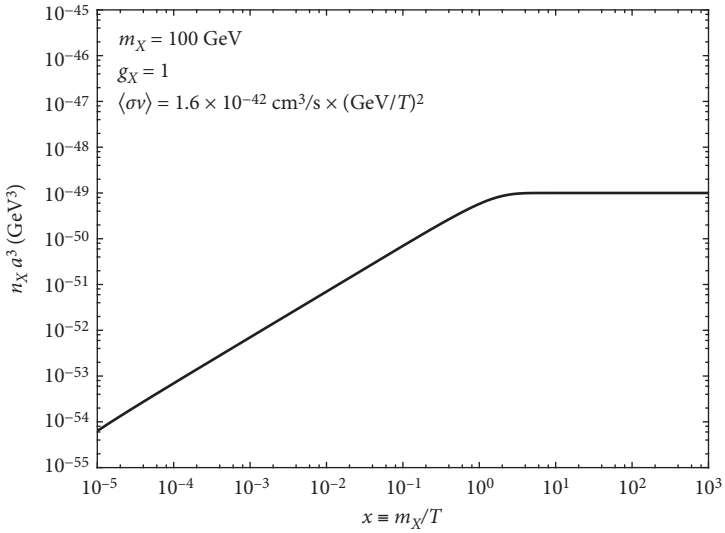


Figure 9.2 The comoving number density of a dark matter candidate undergoing freeze in, as a function of temperature. The case shown in that of a stable scalar with $g_X = 1$, $m_X = 100\,\mathrm{GeV}$, and $\langle\sigma\, v\rangle = (1.4 \times 10^{-25})/T^2 \approx 1.6 \times 10^{-42}\,\mathrm{cm}^3/\mathrm{s} \times (\mathrm{GeV}/T)^2$.

density as the universe expands:

$$\frac{\rho_Y}{\rho_{\mathrm{SM}}} = \begin{cases} \left(\dfrac{30}{\pi^2 g_\star(T) T_i^4}\right) \rho_{Y,i} \left(\dfrac{g_{\star,S}(T)}{g_{\star,S}(T_i)}\right)^{4/3} \propto \dfrac{g_{\star,S}^{4/3}(T)}{g_\star(T)}, & \rho_Y \gg n_Y m_Y, \\ \left(\dfrac{30}{\pi^2 g_\star(T) T_i^4}\right) n_{Y,i}\, m_Y \left(\dfrac{T_i}{T}\right)\left(\dfrac{g_{\star,S}(T)}{g_{\star,S}(T_i)}\right) \propto \dfrac{g_{\star,S}(T)}{g_\star(T)} \dfrac{1}{T}, & \rho_Y \approx n_Y m_Y. \end{cases} \tag{9.54}$$

Up to factors of $g_\star$ and $g_{\star,S}$, $\rho_Y/\rho_{\mathrm{SM}}$ remains constant until the Y population becomes nonrelativistic, after which it grows as T^{-1}. This implies that if Y is sufficiently long-lived, its fractional contribution to the total energy density could potentially grow by a large factor before it decays. In fact, even if the initial abundance of these particles is quite small, the Y population could easily come to dominate the overall energy density of the early universe.

Next, let's postulate that the heavy particles decay with a lifetime τ_Y, and that most of these decays result in the production of Standard Model particles. In addition, let's suppose that a small fraction of these decays produce a pair of very feebly interacting and highly decoupled dark matter particles, $Y \to XX$. Once the Y population has become nonrelativistic, the abundances of these heavy particles, the dark matter, and the energy density of the Standard Model bath evolve as follows:

$$\frac{d\rho_Y}{dt} = -3H\rho_Y - \frac{\rho_Y}{\tau_Y}, \tag{9.55}$$

$$\frac{d\rho_X}{dt} \approx -3Hn_X m_X - 4H(\rho_X - n_X m_X) + BR(Y \to XX)\,\frac{\rho_Y}{\tau_Y},$$

$$\frac{dn_X}{dt} = -3Hn_X + 2\,BR(Y \to XX)\,\frac{n_Y}{\tau_Y},$$
$$\frac{d\rho_{SM}}{dt} = -4H\rho_{\rm SM} + BR(Y \to {\rm SM})\,\frac{\rho_Y}{\tau_Y},$$

where $BR(Y \to XX)$ and $BR(Y \to {\rm SM})$ denote the branching fractions for the decays of the Y to dark matter and to Standard Model species, and the Hubble rate is calculated using the total energy density, $\rho_{\rm tot} = \rho_Y + \rho_X + \rho_{\rm SM}$. Note that the second line of this expression is only an approximation, although an adequate one for our purposes (see Problem 9.12).

In the case that the Y population comes to dominate the total energy density before decaying, these decays will produce a fresh bath of Standard Model particles and dark matter. Treating the decays as instantaneous (all occurring at $t = \tau_Y$), it is straightforward to estimate the temperature of the resulting bath by equating the energy densities immediately before and after the decays. In a matter-dominated universe, the energy density at a time $t = \tau_Y$ is given by $\rho = M_{\rm Pl}^2/6\pi\tau_Y^2$ (see Eq. 2.52). After the decays, the energy density is the usual $\rho = \pi^2 g_\star(T) T^4/30$. Setting these equal, we find that the decays produce a bath of particles with a temperature that is given by

$$T = \left(\frac{5M_{\rm Pl}^2}{\pi^3 g_\star(T)\,\tau_Y^2}\right)^{1/4} \tag{9.56}$$
$$\sim 0.5\,{\rm GeV} \times \left(\frac{60}{g_\star(T)}\right)^{1/4} \left(\frac{10^{-6}\,{\rm s}}{\tau_Y}\right)^{1/2}.$$

If we do this calculation more carefully, no longer assuming that the decays occur instantaneously, but instead proceed according to the rates described in Eq. 9.55, we find that these decaying particles never actually cause the temperature of the universe to increase, but only slow the rate at which the expanding universe cools. In any case, out-of-equilibrium decays can result in the temperature of the universe being much hotter than it would have been otherwise. In the top frame of Figure 9.3, I show how the energy densities of the heavy particles, the Standard Model bath, and the dark matter each evolve in an example in which $\tau_Y = 10^{-6}$ s, $m_Y = 10^6$ GeV, and $m_X = 10^3$ GeV. To obtain the measured abundance of dark matter in this example, we need a branching fraction of $BR(Y \to XX) \approx 2.5 \times 10^{-7}$. The plateau briefly seen in the dark matter energy density curve corresponds to the window of time in which the newly created dark matter particles are still relativistic, and thus evolve as radiation.

Another way in which out-of-equilibrium decays could impact the evolution of the early universe is through the dilution of the abundances of other relics. Consider, for example, a scenario that begins with the dark matter freezing out of equilibrium with the Standard Model bath. Then, at a later time, a species of nonrelativistic and decoupled particles comes to dominate the energy density of the universe before decaying. By injecting energy into the Standard Model bath, these decays reduce the value of ρ_X/ρ_{SM}, suppressing the final dark matter abundance by what could potentially be a large factor. An example of such a scenario is shown in the lower frame of Figure 9.3.

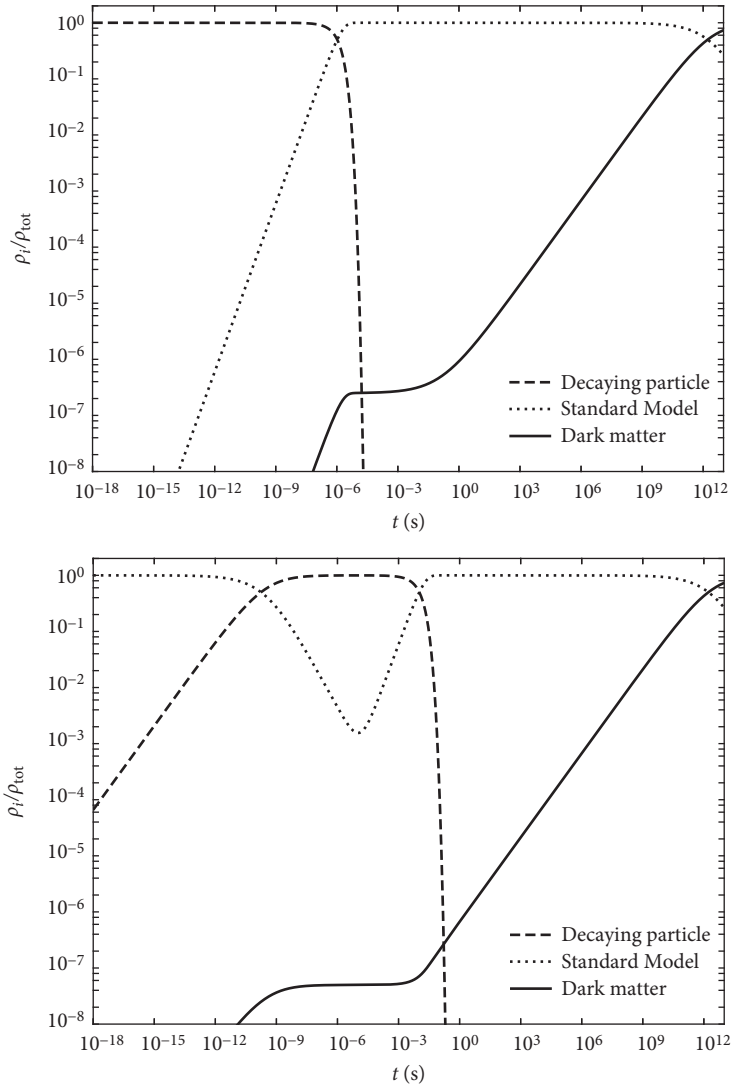


Figure 9.3 The evolution of the energy densities of various components in two scenarios involving the out-of-equilibrium decays of a heavy and decoupled particle species, Y. In the top frame, the Standard Model bath and the dark matter are each produced through these decays, with results shown for the case of $\tau_Y = 10^{-6}$ s, $m_Y = 10^6$ GeV, $m_X = 10^3$ GeV, and $BR(Y \to XX) = 2.5 \times 10^{-7}$. In the lower frame, the heavy particles gradually come to dominate the energy density of the universe before decaying into Standard Model particles (with $\tau_Y = 10^{-2}$ s), diluting the abundance of the (previously generated) dark matter.

In this chapter, we have covered a wide range of scenarios for the production of dark matter in the early universe. Throughout this discussion, I have tried to remain agnostic as to the details of the underlying particle physics theory. We will return to this theme in Chapter 17, focusing on a selection of specific dark matter candidates that can arise within the context of well-motivated particle physics frameworks, including supersymmetric neutralinos, sterile neutrinos, and axions.

Additional reading

The freeze out of a thermal relic is nicely covered in some detail in Kolb and Turner's *The Early Universe*. For a more recent discussion of dark matter and its origin, I recommend Lin's lectures from the TASI summer school [74]. For a historical review of this subject, see Ref. [75].

Problems

9.1 Consider a galaxy containing a total mass of $M = 10^{12}\, M_\odot$ within a radius of 100 kpc and that has a flat rotation curve out to this radius. Assuming spherical symmetry, what is the density of this galaxy as a function of radius?

9.2 Consider a star in an orbit around a galaxy. At what radius does the centrifugal acceleration experienced by the star equal the acceleration scale associated with MOND? This corresponds to the radius at which the effects of MOND would begin to significantly impact the rotational curve of the galaxy.

9.3 Use the expression for MOND in the low-acceleration limit ($F = ma^2/a_0$) to calculate the velocity of a star in a circular orbit around a galaxy as a function of radius. Show that for orbits outside of the majority of the galaxy's mass, the rotation curves are expected to be flat.

9.4 Consider a scalar dark matter candidate with $g_X = 1$ that froze out of equilibrium with the Standard Model bath while still relativistic, when $T_f = 10\,\text{MeV}$. If this particle species makes up 10% of the total dark matter abundance, what must its mass be? What do the constraints on dark radiation from BBN say about this scenario? Do the constraints on dark radiation from the CMB apply in this case? What if the particle had instead decoupled when $T_f = 10\,\text{GeV}$?

9.5 Consider a 5 MeV particle with $g_X = 2$ that froze out of equilibrium as described in Sec. 9.5.2 to become the dark matter. At what temperature must this species have undergone thermal freeze out? Qualitatively, how would you expect this to have impacted the light element abundances produced during BBN?

9.6 Challenge problem: In calculating the thermal average of the annihilation cross section in this chapter, I have made some simplifying approximations. Go through the calculation described in Section 3 of Ref. [21] to obtain:

$$\langle \sigma\, v \rangle = \frac{1}{8 m_X^4\, T K_2^2(m_X/T)} \int_{4m_X^2}^{\infty} \sigma\, s^{1/2} (s - 4m_X^2) K_1(s^{1/2}/T)\, ds, \tag{9.57}$$

where K_i are the modified Bessel functions of the second kind of order, i, and $s = (p_1 + p_2)^2 = E_{\text{CM}}^2$. Then use the fact that $K_1(x) \approx K_2(x) \approx (\pi/2x)^{1/2}\, e^{-x}$ for large values of x to show that this expression reduces to Eq. 9.20 in the nonrelativistic limit.

9.7 Write a code to numerically integrate Eq. 9.28 for the case of $m_X = 100\,\text{GeV}$, $\Gamma_{\text{med}} = 1\ \text{GeV}$, and $\alpha = 10^{-3}$. Plot $\langle \sigma\, v \rangle$ as a function of the mediator mass. Use this code in combination with Eqs. 9.25 and 9.26 to find the thermal relic abundance of this dark matter candidate, again as a function of m_{med}.

9.8 Consider a stable Majorana fermion with a mass of 100 GeV, and an unstable Majorana fermion with a mass of 110 GeV, each of which has a self-annihilation cross section of $10^{-27}\,\text{cm}^3/\text{s}$. For what value of the coannihilation cross section between these two

states will we obtain a thermal relic abundance that is equal to the measured dark matter density?

9.9 Consider a stable Dirac fermion with a mass of 100 GeV and a chemical potential of $\mu_X = -\mu_{\bar{X}} = 50$ GeV. For what annihilation cross section do we obtain a thermal relic abundance equal to the measured dark matter density?

9.10 In deriving Eq. 9.50, I used the approximation $dt \approx -dT/TH$, which is valid up to variations in $g_{\star,S}$. From the relationship between the scale factor and temperature that follows from entropy conservation, derive the exact expression for dt. For $g_{\star,S}(T)$ as predicted in the Standard Model, under what circumstances is this approximation least valid? How much error did this approximation introduce into my calculation of the dark matter abundance from freeze in?

9.11 Consider a dark matter candidate that is produced through freeze in, but with a pair production cross section that is inversely proportional to the fourth power of the temperature, $\langle \sigma v \rangle = (\sigma v)_0 / T^4$. Derive the final abundance of dark matter that is produced in this case. What value of $(\sigma v)_0$ is required to obtain the measured dark matter density? Is this an example of an IR-dominated or UV-dominated freeze-in model?

9.12 Neglecting production through decays, I wrote in Eq. 9.55 that the energy density in the particle species, X, will evolve as

$$\frac{d\rho_X}{dt} \approx -3Hn_X m_X - 4H(\rho_X - n_X m_X).$$

As pointed out in the text, this is only an approximation. In generality, we can write the evolution of the energy density of a particle species in terms of its equation of state, $\rho_X \propto a^{-3(1+w_X)}$ (see Eq. 2.47). Take the derivative of this expression to obtain a form similar to that given in Eq. 9.55 and compare these expressions in the relativistic ($T \gg m$) and nonrelativistic ($T \ll m$) limits. Recall that the equation of state of a species is related to its pressure and energy density, $w \equiv p/\rho$.

9.13 In the Boltzmann equations given in Eq. 9.55, the unstable particles were assumed to be nonrelativistic by the time of their decays. If this is not the case, their decay rate will be reduced by the effects of time dilation. In more generality, the Boltzmann equation describing the evolution of the number density of a decaying particle species, X, can be written as follows:

$$\frac{dn_X}{dt} = -3Hn_X - \langle \Gamma_X \rangle (n_X - n_X^{\rm Eq}). \tag{9.58}$$

The thermally averaged decay width in this expression is given by

$$\langle \Gamma_X \rangle = \Gamma_X \frac{K_1(x)}{K_2(x)}, \tag{9.59}$$

where $K_i(x)$ are the modified Bessel functions of the second kind and $x \equiv m_X/T$ (see Appendix A of Ref. [76] for a derivation of this expression). Show that $\langle \Gamma_X \rangle$ reduces to $\Gamma_X = \tau_X^{-1}$ in the large x limit. Also, in writing Eq. 9.55, why could we neglect the term associated with inverse decays, $dn_X/dt \propto \langle \Gamma_X \rangle n_X^{\rm Eq}$?

9.14 As a variation of dark matter production through out-of-equilibrium decays (as discussed in Sec. 9.7.2), consider a scenario in which the early universe was dominated by black holes, each with a mass of $M_{\rm BH} = 10^8$ grams. Black holes produce

particles through Hawking evaporation, radiating at a temperature of $T \sim 10^{13}\,\text{GeV} \times (\text{g}/M_{\text{BH}})$, and causing them to lose mass at a rate given by

$$\frac{dM_{\text{BH}}}{dt} \approx -\frac{\mathcal{G} g_{\star,H} M_{\text{Pl}}^4}{30720\pi M_{\text{BH}}^2}, \tag{9.60}$$

where $\mathcal{G} \approx 3.8$ and $g_{\star,H} \approx 108$ for $T_{\text{BH}} \gtrsim 100\,\text{GeV}$.

Treating the process of black hole evaporation as instantaneous (occurring at the time at which $M_{\text{BH}} = 0$), calculate the temperature of the Standard Model bath after the black holes have completely evaporated. If 1% of the energy emitted from these black holes is in the form of the mass of dark matter particles with $m_X = 100\,\text{GeV}$, estimate the resulting dark matter density in our universe today. What if these dark matter particles instead had a mass of $m_X = 7 \times 10^{10}\,\text{GeV}$? Hint: See Ref. [77].

10

The Formation of Large-Scale Structure

Over the past four chapters, we have focused on events that took place during the first few hundred thousand years after the Big Bang. During those early times, matter and energy were distributed in a very uniform way, featuring only very small variations in density, $\delta\rho/\rho \sim 10^{-5}$. And while our current universe remains approximately homogeneous on scales larger than ~ 100 Mpc, it is not at all uniform on smaller scales. Over the course of cosmic history, much of the baryonic matter became gathered together into galaxies, as well as compact objects such as stars and planets. This was made possible through the baryons' electromagnetic interactions, which allowed them to cool and lose angular momentum. Dark matter, in contrast, does not experience such interactions, and thus has not been able to collapse as completely as baryons. Instead, the force of gravity has gradually amplified the small density perturbations that were present in the early universe, ultimately leading to the formation of the gravitationally bound dark matter halos that have served as the scaffolding on which the large-scale structure of our universe has been built. In this chapter, I will review this process of structure formation and discuss what the large-scale distribution of matter in our universe can tell us about the nature of dark matter, and about our universe more generally.

10.1 Gravitational collapse

Consider a large volume of space with an approximately uniform density, $\rho(\vec{x}) \approx \bar{\rho}$. Within that volume, imagine that there is a smaller spherical region of radius R with a density that is slightly higher or lower than the average value, $\rho = \bar{\rho}\,(1+\delta)$. Gravity will act on the matter within this region, pulling or pushing it toward its center of mass, and thus causing the region to collapse or expand, depending on whether δ is positive or negative. Using Newton's second law, we can write

$$\frac{d^2R}{dt^2} = -\frac{G\,\Delta m}{R^2} = -\frac{4\pi\,G\bar{\rho}R\delta}{3}, \tag{10.1}$$

where $\Delta m = (\rho - \bar{\rho})(4\pi R^3/3)$ is the excess mass contained within the spherical region above that associated with the mean density. We can relate R and δ using the fact that the total amount of mass contained within the radius, R, remains constant as the region collapses

and increases in density (or expands and decreases in density):

$$M = \frac{4\pi}{3} R^3 \, \bar{\rho} \, [1 + \delta], \tag{10.2}$$

and thus

$$\begin{aligned} R &= R_0 \, [1 + \delta]^{-1/3} \\ &\approx R_0 \left[1 - \frac{\delta}{3} \right], \end{aligned} \tag{10.3}$$

where $R_0 \equiv (3M/4\pi\bar{\rho})^{1/3}$ is a constant. In the second line, I have performed a Taylor series expansion in the $\delta \ll 1$ limit.

Taking the second derivative of R with respect to time, we arrive at

$$\begin{aligned} \frac{d^2R}{dt^2} &= -\frac{R_0}{3} \frac{d^2\delta}{dt^2} \\ &\approx -\frac{R}{3} \frac{d^2\delta}{dt^2}. \end{aligned} \tag{10.4}$$

Combining this with Eq. 10.1, we have

$$\frac{d^2\delta}{dt^2} = 4\pi G \bar{\rho} \, \delta, \tag{10.5}$$

whose general solution is given by

$$\delta = A e^{t/\tau_c} + B e^{-t/\tau_c}, \tag{10.6}$$

where A and B are constants determined by the initial conditions, and τ_c is the timescale for collapse:

$$\begin{aligned} \tau_c &= (4\pi G \bar{\rho})^{-1/2} \\ &\approx 7.5 \times 10^4 \, \mathrm{yr} \times \left(\frac{2.1 \times 10^{-22} \, \mathrm{kg/cm^3}}{\bar{\rho}} \right)^{1/2}. \end{aligned} \tag{10.7}$$

In this last line, I have scaled this result in terms of the total energy density at the time of matter-radiation equality.

At face value, this calculation would appear to show that the force of gravity will cause denser-than-average regions to collapse exponentially, forming increasingly dense structures without limit (or at least until $\delta \sim 1$, at which point our approximations break down). Furthermore, this should occur on all length scales. After all, the timescale for collapse is independent of the size of the overdense region, R, and only depends on the average density, $\bar{\rho}$. If this had actually occurred, the energy density of the universe would have rapidly become very clumpy ($\delta \gtrsim 1$), in stark disagreement with observations.

In practice, there are two important factors which work to prevent the exponential collapse of overdense regions. First, pressure pushes back against the effects of gravity,

especially in the case of photons and baryonic matter. Second, the expansion of the universe suppresses this kind of exponential growth. Notice that the Hubble time is similar in magnitude to the timescale for collapse, $H^{-1} \sim \tau_c$, and that both of these quantities scale as $\bar{\rho}^{-1/2}$, leading us to expect that Hubble expansion will meaningfully impact the growth of density perturbations throughout all periods of cosmic history. In the following two sections, we will explore the impact of each of these factors on the formation of large-scale structure in our universe.

10.2 Pressure and the Jeans length

Even a casual look at Eqs. 10.6 and 10.7 will reveal that this solution cannot represent the whole story of gravitational collapse. Just consider the cup of coffee that is sitting on the table before me at this very moment. With a density of $\rho \approx 1\ \mathrm{g/cm^3}$, these expressions would suggest that gravity should cause this smooth and homogeneous fluid to form dense clumps on a timescale of about $\tau_c \approx 18$ minutes. A similar calculation will show that the air in my office should collapse into dense clumps after only 10 hours or so. Obviously, this is not taking place. The reason why is pressure.

For the case of an ideal gas, we can write the pressure as follows:

$$p \approx n k_B T = \frac{\rho k_B T}{m}, \tag{10.8}$$

where n is the number density of particles of mass, m. For the sake of familiarity, I have included in this expression the Boltzmann factor, k_B, which is ordinarily suppressed in the system of natural units. I will omit this symbol in what follows.

From the pressure, we can calculate the speed of sound in the (nonrelativistic) ideal gas:

$$c_s^2 = \frac{dp}{d\rho} = c^2\,\frac{T}{m}. \tag{10.9}$$

In other words, the equation of state of the gas is given by $w = T/m$.

From this, we see that across a region of size R, the effects of pressure are able to react to compression on a timescale given by $\tau_p \sim R/c_s \approx R\,m^{1/2}/T^{1/2}$. Pressure will be able to meaningfully prevent gravitational collapse only if $\tau_p \lesssim \tau_c$. In the case of the air in my office (or the coffee in my cup), the time for sound to propagate across the region is much, much shorter than the timescale for collapse, $\tau_p \ll \tau_c$. As a result, hydrostatic equilibrium is maintained and pressure prevents any overdense regions from collapsing under the influence of gravity. Putting this another way, gravitational collapse is suppressed on all length scales smaller than what is known as the Jeans length, λ_J, which is defined as follows:

$$\begin{aligned} \lambda_J &\equiv 2\pi\, c_s \tau_c \\ &= \left(\frac{\pi\ T}{G\bar{\rho}\ m}\right)^{1/2} \\ &= \left(\frac{w\pi}{G\bar{\rho}}\right)^{1/2}. \end{aligned} \tag{10.10}$$

Alternatively, we can write the Jeans length in terms of the Hubble rate:

$$\lambda_J = \left(\frac{w\pi}{G\bar{\rho}}\right)^{1/2} = 2\pi \left(\frac{2w}{3}\right)^{1/2} \frac{1}{H}. \tag{10.11}$$

During the era of radiation domination, the equation of state was $w \approx 1/3$, and this prevented photons and baryons from experiencing gravitational collapse on all scales smaller than $\lambda_J \sim 3\, H^{-1}$. The only subhorizon density perturbations in the photon-baryon fluid whose evolution was not suppressed were sound waves (i.e., acoustic oscillations), as discussed in Chapter 7. These acoustic oscillations were stable, however, and did not collapse under the influence of gravity or otherwise lead to the formation of gravitationally bound structures.

In order for gravity to produce structures on scales smaller than the Hubble length, there must have been a component of the energy density that had a significantly smaller equation of state, allowing for a lower sound speed and Jeans length. In our universe, this was accomplished through the presence of nonrelativistic matter, both baryonic and dark, which provided precisely what was needed to allow for the generation of gravitationally collapsed structures.

Prior to recombination, the photons and the baryonic matter were tightly coupled to one another. As a consequence, the photons provided pressure that prevented the baryons from collapsing under the influence of gravity. After decoupling, however, the baryons were liberated from this pressure, making it possible for them to begin to form gravitationally collapsed structures. The Jeans length of the baryonic matter during the era of matter domination and after decoupling was given by

$$\begin{aligned} \lambda_J &= \left(\frac{\pi\, T_b}{G\bar{\rho}\, m}\right)^{1/2} \\ &= \left(\frac{\pi (1+z)^2 T_{\rm dc}/(1+z_{\rm dc})^2}{G\,\Omega_M \rho_{\rm crit} (1+z)^3\, m}\right)^{1/2} \\ &\approx 5.6 \times 10^{19}\ {\rm cm} \times \left(\frac{1090}{1+z}\right)^{1/2}, \end{aligned} \tag{10.12}$$

or in terms of the photon temperature, T:

$$\begin{aligned} \lambda_J &= \left(\frac{\pi\, T_{\rm CMB}^3}{G\,\Omega_M \rho_{\rm crit} m\, T\, T_{\rm dc}}\right)^{1/2} \\ &\approx 5.6 \times 10^{19}\ {\rm cm} \times \left(\frac{0.26\ {\rm eV}}{T}\right)^{1/2}, \end{aligned} \tag{10.13}$$

where $z_{\rm dc} \approx 1{,}090$ and $T_{\rm dc} = T_{\rm CMB}\,(1+z_{\rm dc}) \approx 0.26\ {\rm eV}$ are the redshift and temperature at the time of photon decoupling. In deriving these expressions, I have made use of the fact that the temperature of a population of decoupled, nonrelativistic particles evolves as $T_b \propto (1+z)^2$ as space expands, while the photon temperature instead evolves as $T \propto (1+z)$ (see Sec. 4.6).

From these expressions, we can calculate the baryonic Jeans mass, which is defined as the total mass of baryons enclosed within a sphere of radius λ_J:

$$M_J = \frac{4\pi\lambda_J^3}{3}\,\rho_b \approx 2\times 10^5\,M_\odot \times \left(\frac{1+z}{1090}\right)^{3/2}, \tag{10.14}$$

where $\rho_b = \rho_b(z=0)\,(1+z)^3$.

Since the time of decoupling, baryons could undergo gravitational collapse on all mass scales larger than M_J. When evaluated at the time of decoupling, M_J is comparable to the baryonic masses of the smallest dwarf galaxies observed in our universe. The growth of smaller baryonic structures, in contrast, remained suppressed by the effects of pressure.

In contrast to baryons, the dark matter decoupled from the photons much earlier, giving them more time to cool and thus enabling them to form smaller structures. For a dark matter particle that underwent kinetic decoupling at a temperature $T_{\rm kd}$, the Jeans length of this substance during the radiation-dominated era is given by

$$\begin{aligned}\lambda_J &= \left(\frac{\pi\,T_{\rm DM}}{G\bar{\rho}\,m_{\rm DM}}\right)^{1/2} \\ &= \left(\frac{\pi\,T_{\rm kd}}{G\,\Omega_{\rm rad}\,\rho_{\rm crit}(1+z)^4\,m_{\rm DM}}\,\frac{(1+z)^2}{(1+z_{\rm kd})^2}\right)^{1/2} \\ &= \left(\frac{\pi\,T_{\rm CMB}^4\,g_{\star,S}^{4/3}(T_{\rm CMB})}{G\,\Omega_{\rm rad}\,\rho_{\rm crit} m_{\rm DM}\,T^2\,T_{\rm kd}\,g_{\star,S}^{2/3}(T)\,g_{\star,S}^{2/3}(T_{\rm kd})}\right)^{1/2} \\ &\approx 1.3\times 10^9\,{\rm cm}\times\left(\frac{1\,{\rm MeV}}{T}\right)\left(\frac{1\,{\rm MeV}}{T_{\rm kd}}\right)^{1/2}\left(\frac{100\,{\rm GeV}}{m_{\rm DM}}\right)^{1/2}\left(\frac{g_{\star,S}^2(T_{\rm CMB})}{g_{\star,S}(T)\,g_{\star,S}(T_{\rm kd})}\right)^{1/3},\end{aligned} \tag{10.15}$$

where in the second line I have expressed the mean density in terms of the modern values of $\Omega_{\rm rad}$ and $\rho_{\rm crit}$, $\bar{\rho} = \Omega_{\rm rad}\,\rho_{\rm crit}\,(1+z)^4$. Recall that being nonrelativistic and decoupled, the temperature of the dark matter evolves as $T_{\rm DM} = T_{\rm kd}\,(1+z)^2/(1+z_{\rm kd})^2$. In deriving the third line, I applied entropy conservation to express powers of $(1+z)$ and $(1+z_{\rm kd})$ in terms of ratios of temperatures and $g_{\star,S}$.

A region with a radius equal to the Jeans length encloses the following mass of dark matter:

$$\begin{aligned}M_J &= \frac{4\pi}{3}\lambda_J^3\,\rho_{\rm DM} \\ &= \frac{4\pi^{5/2}\,T_{\rm CMB}^6\,g_{\star,S}^2(T_{\rm CMB})\,\Omega_{\rm DM}(1+z)^3}{3G^{3/2}\,\Omega_{\rm rad}^{3/2}\,\rho_{\rm crit}^{1/2} m_{\rm DM}^{3/2}\,T^3\,T_{\rm kd}^{3/2}\,g_{\star,S}(T)\,g_{\star,S}(T_{\rm kd})} \\ &= \frac{4\pi^{5/2}\,\Omega_{\rm DM}\,T_{\rm CMB}^3\,g_{\star,S}(T_{\rm CMB})}{3\rho_{\rm crit}^{1/2}\,G^{3/2}\Omega_{\rm rad}^{3/2} m_{\rm DM}^{3/2}\,T_{\rm kd}^{3/2}\,g_{\star,S}(T_{\rm kd})} \\ &\approx 3\times 10^{-7}\,M_\odot\times\left(\frac{100\,{\rm GeV}}{m_{\rm DM}}\right)^{3/2}\left(\frac{1\,{\rm MeV}}{T_{\rm kd}}\right)^{3/2}\left(\frac{10.7}{g_{\star,S}(T_{\rm kd})}\right),\end{aligned} \tag{10.16}$$

where in the second line I have used $\rho_{\rm DM} = \Omega_{\rm DM}\,\rho_{\rm crit}\,(1+z)^3$, and in the third line I have used entropy conservation to equate $(1+z)^3 = T^3 g_{\star,S}(T)/T^3_{\rm CMB}\,g_{\star,S}(T_{\rm CMB})$.

Although pressure prevents baryonic matter from undergoing gravitational collapse until after recombination, density perturbations can begin to grow among the dark matter particles at much earlier times, as soon as they have kinetically decoupled from the Standard Model bath. As a consequence, the dark matter can form much smaller structures, depending on the mass of the particles that make up this substance and on the temperature at which they underwent kinetic decoupling. In many WIMP models, for example, the dark matter can collapse to form structures that are as small as $M_J \sim 10^{-8} - 10^{-3}\,M_\odot$.

When the effects of pressure are taken into account, the evolution of density perturbations as described in Eq. 10.5 is modified as follows:

$$\frac{d^2\delta}{dt^2} + \left[\left(\frac{\lambda_J}{\lambda_p}\right)^2 - 1\right] 4\pi G\bar{\rho}\,\delta = 0, \tag{10.17}$$

where λ_p is the proper size of the perturbed region (more precisely, λ_p is the proper wavelength of a mode in the spectrum of density perturbations, as we will discuss in Sec. 10.4). In the $\lambda_p \gg \lambda_J$ limit, the behavior found in Eq. 10.5 is recovered, and we once again predict the exponential growth of any density perturbations, at least when neglecting the effects of Hubble expansion. In the opposite limit, we find a very different result. In particular, on length scales much smaller than the Jeans length, $\lambda_p \ll \lambda_J$, the solution to this equation is given by $\delta \propto \cos(\omega t)$, where $\omega = (\lambda_J/\lambda_p)(4\pi G\bar{\rho})^{1/2}$. This solution describes the propagation of acoustic waves, which oscillate rather than grow in amplitude.

Lastly, note that Eq. 10.17 is limited to cases in which the energy density consists of only a single component. We can straightforwardly generalize this calculation to consider a universe that is filled with multiple substances (such as photons, baryons, dark matter, etc.):

$$\frac{d^2\delta_i}{dt^2} + \left(\frac{\lambda_{J,i}}{\lambda_p}\right)^2 4\pi G\bar{\rho}\,\delta_i - 4\pi G\bar{\rho}\sum_j \epsilon_j\delta_j = 0, \tag{10.18}$$

where $\bar{\rho}$ is the total mean density, δ_i and $\lambda_{J,i}$ are the density perturbation and Jeans length of the ith substance, and ϵ_j is the fraction of the total energy density that is made up of substance, j. Notice that while the effects of pressure only impact the substances that supply that pressure (or that are coupled to substances that supply the pressure), all of the forms of energy contribute to the gravitational potential.

10.3 The evolution of structure in an expanding universe

So far in this chapter, we have largely ignored the fact that our universe is expanding. As we saw in Sec. 10.1, however, the timescale for gravitational collapse has been comparable to that of Hubble expansion throughout cosmic history, $\tau_c \sim H^{-1}$, and thus cannot be safely ignored. The expansion of space has the effect of suppressing the growth of structure, preventing the exponential behavior that we would have predicted in a static universe.

We will begin this calculation with a Newtonian treatment, which is a valid approximation on scales much smaller than that of the horizon. In the Newtonian regime, density perturbations in a background of expanding space evolve according to the following

equation:

$$\frac{d^2\delta_i}{dt^2} + 2H\frac{d\delta_i}{dt} + \left(\frac{\lambda_{J,i}}{\lambda_p}\right)^2 4\pi G\bar{\rho}\,\delta_i - 4\pi G\bar{\rho}\sum_j \epsilon_j\delta_j = 0. \tag{10.19}$$

Notice that this is the same expression as in Eq. 10.18, but with an additional term that is proportional to the Hubble rate. This term acts like "Hubble friction," and serves to slow the growth of density perturbations.[1]

During the era of radiation domination, the large Jeans length prevented any density perturbations in the photon-baryon fluid from growing on subhorizon scales. The dark matter, however, was impacted by pressure only on very small scales, making it possible for their density perturbations to grow during this era. During radiation domination ($\rho_{\rm DM} \ll \rho_{\rm tot}$, $H = 1/2t$), the solution to Eq. 10.19 for the dark matter is given as follows:

$$\delta_{\rm DM} = A + B\ln t, \tag{10.20}$$

where A and B are constants determined by the initial conditions. From this, we learn that during periods of radiation domination, subhorizon density perturbations grow quite slowly in an expanding universe, no faster than logarithmically.

During the matter-dominated era, the fourth term in Eq. 10.19 cannot be neglected. Using the fact that $H = 2/3t$ during this period, we can describe the evolution of the subhorizon density perturbations on all scales larger than the Jeans length as follows:

$$\frac{d^2\delta_{\rm DM}}{dt^2} + \frac{4}{3t}\frac{d\delta_{DM}}{dt} - \frac{2}{3t^2}\delta_{DM} = 0. \tag{10.21}$$

The solution to this equation is given by $\delta_{\rm DM} = At^{2/3} + Bt^{-1}$, where A and B are once again determined by initial conditions. In practice, the t^{-1} piece of this expression quickly disappears, leaving the perturbations to grow as $\delta_{\rm DM} \propto t^{2/3}$, which during matter domination is equivalent to $\delta_{\rm DM} \propto a \propto (1+z)^{-1}$. Since the era of matter domination lasted from $(1+z) \approx 3{,}400$ until $(1+z) \approx 1.3$, density perturbations were able to grow by a factor of $\sim 3 \times 10^3$ during this period of time, transforming the nearly uniform distribution of matter that existed during the era of recombination into the collection of dense galaxies and galaxy clusters that are found in our universe today.

As matter was diluted and dark energy began to dominate the energy density of our universe, the growth of density perturbations was once again suppressed. In the dark energy-dominated limit, the fourth term in Eq. 10.19 is negligible, as it was in the radiation-dominated case. Furthermore, $H = (8\pi G\rho_\Lambda/3)^{1/2}$ is a constant in this limit, leading to a solution of the form $\delta_{DM} = A + Be^{-2Ht}$, which quickly evolves to a nearly constant value of δ_{DM}.

In the present era, there are significant densities of both dark energy and matter, and the dark energy has not yet brought the growth of density perturbations to an end. The solution to Eq. 10.19 for a flat universe that contains both matter and dark energy is approximately given by $d\ln\delta_{DM}/d\ln a \approx (\rho_{DM}/\rho_{\rm tot})^{0.55}$ [80, 81]. In other words, the dependance on the

[1] In an expanding universe, the baryonic Jeans length evolves with time, and the resulting suppression of the growth of density perturbations depends on a time-averaged value of this quantity. The effective mass below which baryonic pressure significantly suppresses the growth of structure is known as the "filtering mass" [78, 79].

scale factor gradually evolves from that exhibited during matter domination, $\delta_{DM} \propto a$, to the constant behavior that is predicted in the dark energy-dominated limit, $\delta_{DM} = \text{constant}$.

To describe the evolution of any density perturbations that exist on scales larger than the size of the horizon, a general relativistic treatment is required. To see one the most important results of such an analysis, consider a spatially flat universe with a Hubble rate that is given by $H = (8\pi G\rho/3)^{1/2}$. Next consider a perturbed region of that universe, with the same rate of Hubble expansion, but containing a greater density, $\rho' > \rho$, and thus exhibiting a positive curvature. Using the first Friedmann equation, we can write

$$H^2 = \frac{8\pi G\rho}{3} = \frac{8\pi G\rho'}{3} - \frac{\kappa}{a^2}, \tag{10.22}$$

which we can then use to find the fractional difference between the two densities:

$$\delta \equiv \frac{\rho' - \rho}{\rho} = \frac{3\kappa}{8\pi G a^2 \rho}. \tag{10.23}$$

Combining this with the fact that $\rho \propto a^{-4}$ during radiation domination and $\rho \propto a^{-3}$ during matter domination, we find that

$$\delta \propto \frac{1}{a^2\rho} \propto \begin{cases} a^2 \propto t & \text{(radiation domination)}, \\ a \propto t^{2/3} & \text{(matter domination)}. \end{cases} \tag{10.24}$$

In the matter-dominated case, we have obtained the same result that was found in our Newtonian analysis described above. During radiation domination, however, density perturbations on superhorizon scales grew much more quickly than was suggested by our Newtonian treatment. In fact, these perturbations grew even more rapidly during radiation domination than they did during the matter-dominated era.

Let's pause and summarize the main conclusions of the past two sections. Subhorizon density perturbations began to grow at a significant rate in our universe, $\delta \propto a$, around the time of matter-radiation equality, and then only among the dark matter particles. The baryons remained tightly coupled to the photons, and thus experienced far too much pressure to undergo gravitational collapse. The density perturbations in the dark matter grew across a wide range of scales. After photon decoupling, the baryons began to collapse on scales larger than roughly the mass of the smallest observed galaxies. On superhorizon scales, density perturbations began to evolve very early, growing particularly rapidly during the radiation-dominated era.

10.4 The matter power spectrum

In Chapter 7, we characterized the statistical properties of the CMB's temperature (and polarization) anisotropies in terms of its angular power spectrum. Strictly speaking, this angular power spectrum does not contain *all* of the information that is encoded in the CMB's distribution, but as long as the anisotropies are drawn from a distribution that is both Gaussian and symmetric with respect to rotations, it will fully encapsulate all of the information that we can hope to extract regarding the statistical distribution of the CMB as a function of angular scale. Thus, up to possible departures from Gaussianity, the CMB's

angular power spectrum represents the totality of this signal's information that can be used to measure and constrain our universe's cosmological parameters.

In a similar spirit, cosmologists are not generally interested in the density of matter that happens to be present at any particular location, but rather in the statistical properties of our universe's matter distribution. To this end, we can express the distribution of density perturbations in terms of a Fourier expansion:

$$\delta(\vec{x}) = \sum_{l,m,n=-\infty}^{\infty} \delta_{\vec{k}}\, e^{-i\vec{k}\cdot\vec{x}}, \tag{10.25}$$

where $\vec{k}$ is the three-dimensional wavenumber. Here, I have imposed periodic boundary conditions on a cube of length, L, ensuring that $\vec{k} \equiv (2\pi l/L, 2\pi m/L, 2\pi n/L)$. In the large L limit, we can write this as an integral:

$$\delta(\vec{x}) = \frac{V}{(2\pi)^3} \int_V \delta_{\vec{k}}\, e^{-i\vec{k}\cdot\vec{x}}\, d^3k, \tag{10.26}$$

where $V = L^3$, and the coefficients, $\delta_{\vec{k}}$, are given by

$$\delta_{\vec{k}} = \frac{1}{V} \int_V \delta(\vec{x})\, e^{i\vec{k}\cdot\vec{x}}\, d^3k. \tag{10.27}$$

Through this exercise, we have written the distribution of the density variations, $\delta(\vec{x})$, in terms of an infinite series of sine waves of comoving wavenumber, $\vec{k}$, and comoving wavelength, $\lambda = 2\pi/|\vec{k}|$. This approach is useful in part because each of these modes evolves independently of one another, so long as the density variations are small, $|\delta_{\vec{k}}| \ll 1$. More specifically, for each mode, $\vec{k}$, the magnitude of $\delta_{\vec{k}}$ evolves as described in the previous section, while the complex phase of each $\delta_{\vec{k}}$ remains constant.

The matter power spectrum is defined as the mean square amplitude of the Fourier coefficients:

$$P(k) \equiv \langle |\delta_{\vec{k}}|^2 \rangle, \tag{10.28}$$

where the average is taken over a large volume of space, which is statistically equivalent to taking the average over a large ensemble of realizations. Note that because of our assumption of isotropy, we can write the power spectrum simply as $P(k)$ rather than as $P(\vec{k})$.

So long as 1) the phases of $\delta_{\vec{k}}$ are uncorrelated with one another, 2) δ is a Gaussian field, and 3) the matter distribution is homogeneous and isotropic, the statistical properties of this field are fully encapsulated in the matter power spectrum, $P(k)$. In particular, the values of δ found throughout space follow a Gaussian distribution centered around zero and with a variance given by

$$\begin{aligned} \sigma_\delta^2 &= \frac{V}{(2\pi)^3} \int P(k)\, d^3k \\ &= \frac{V}{2\pi^2} \int_0^\infty P(k)\, k^2\, dk. \end{aligned} \tag{10.29}$$

The statistical properties of the matter power spectrum can also be described in terms of the N-point correlation functions. For example, the 2-point correlation function, $\xi(r)$,

describes the average of all possible products of two δs separated by a distance r:

$$\xi(r) = \langle \delta(x)\delta(x+r) \rangle. \tag{10.30}$$

This is also sometimes referred to as the autocorrelation function. Note that the correlation functions are the Fourier conjugates of the power spectrum, and thus contain the same underlying information.

The matter power spectrum of our universe depends not only on how density perturbations evolve over time, but also on the initial conditions. These initial conditions can be characterized in terms of the spectrum of primordial density perturbations, which is often parameterized as a power-law, $P(k) \propto k^{n_s}$. As discussed in Sec. 7.5.1, $n_s = 1$ corresponds to the scale-invariant case, known as the Harisson-Peebles-Zel'dovich spectrum. As we have explored over the past few sections, the spectrum of density perturbations does not remain in its primordial state, but evolves as the effects of gravity, pressure, and expansion each act on the matter and energy that occupy our universe. The collective impact of these factors is encapsulated in what is called the transfer function, T_k, such that the matter power spectrum today is given by

$$P(k) \propto k^{n_s}\, T_k^2. \tag{10.31}$$

Recall that each mode evolves independently, allowing us to determine the final matter power spectrum simply by multiplying the primordial spectrum by the square of the transfer function at each value of k. Under the assumption that the dark matter is cold, the transfer function depends on only four parameters: h, Ω_b, $\Omega_{\rm DM}$, and an overall normalization. This normalization is conventionally written in terms of σ_8, defined as the amplitude of the linear power spectrum on a scale of $8\,h^{-1}$ Mpc (see Eq. 10.29). Measurements of the matter power spectrum can thus be used to constrain the values of these parameters, as well as that of n_s. They can also be used to test whether the dark matter is cold, warm, or hot, as we will discuss in Sec. 10.6.

10.5 From density perturbations to dark matter halos

So far in this chapter, we have assumed that the density field is approximately uniform, featuring only small perturbations around the average value ($\delta \ll 1$). Over the course of cosmic history, however, most of the matter in our universe has fallen into dense, self-gravitating clumps, leaving the distribution of matter highly nonuniform on scales smaller than $\sim \mathcal{O}(10^2$ Mpc). Once these density perturbations become large, $\delta \gtrsim 1$, linear theory breaks down and numerical techniques must instead be used to reliably predict how these dense clumps will change over time. While the details of nonlinear structure formation are beyond the scope of this chapter, we will briefly review some of the most important aspects of this subject here. By the end of this section, we will be able to calculate the approximate distribution of gravitationally bound halos that form through the process of gravitational collapse.

Consider a large and approximately uniform volume of space that contains a smaller region with a higher than average density of matter. As we did in Eq. 10.22, we can treat this density perturbation as a region of positive spatial curvature embedded within an otherwise flat universe. The radius of the overdense region thus evolves in the same way that the scale factor of a closed universe does, as described in Sec. 2.7.2. More specifically, the evolution

of the scale factor associated with the overdense region is described as (see Chapter 2, Problem 2.15)

$$a = \frac{4\pi G\rho_0}{3}(1 - \cos\eta), \tag{10.32}$$

where η is related to time according to the following:

$$t = \frac{4\pi G\rho_0}{3}(\eta - \sin\eta). \tag{10.33}$$

For small t, these expressions can be combined and written as follows (see Problem 10.5):

$$a \approx \frac{2\pi G\rho_0}{3}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right]. \tag{10.34}$$

Thus, for $t \ll G\rho_0$, the size of the region grows at the standard rate predicted during matter domination, $a \propto t^{2/3}$, with the radius of the overdensity simply expanding with the Hubble flow. Since the density within the spherical region evolves as $\rho \propto a^{-3} \propto 1 + \delta$, we can also write

$$1 + \delta_{\rm linear} = \left[1 - \frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right]^{-3}, \tag{10.35}$$

where the subscript "linear" denotes the fact that this result applies only to the case of $t \ll G\rho_0$, at which time $\delta \ll 1$. Taking $(1 + \delta_{\rm linear})^{-1/3} \approx 1 - \delta_{\rm linear}/3$, we arrive at the following expression for the evolution of the density perturbation in the linear regime:

$$\begin{aligned} 1 - \frac{\delta_{\rm linear}}{3} &\approx 1 - \frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3} \\ \longrightarrow\ \delta_{\rm linear} &\approx \frac{3}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}. \end{aligned} \tag{10.36}$$

Notice that this agrees with the result found in Sec. 10.3 for the case of matter domination, $\delta \propto t^{2/3}$.

From Eqs. 10.32 and 10.33, we see that the region expands to a maximum size and starts to collapse when $\eta = \pi$, at which point $a_{\rm max} = 8\pi G\rho_0/3$ and $t_{\rm max} = 4\pi^2 G\rho_0/3$. At that time, the density contrast is given by

$$\begin{aligned} \left.\frac{\rho}{\rho_{\rm bg}}\right|_{t_{\rm max}} &= \frac{\rho_0\, a_{\rm max}^{-3}}{\rho_{\rm crit,0}\, a^{-3}} \\ &= \frac{\rho_0}{\rho_{\rm crit,0}}\left[\left(\frac{3}{2}\right)^{2/3} t_{\rm max}^{2/3}\left(\frac{8\pi G\rho_{\rm crit,0}}{3}\right)^{1/3}\right]^3\left[\frac{8\pi G\rho_0}{3}\right]^{-3} \\ &= \frac{9\pi^2}{16}, \end{aligned} \tag{10.37}$$

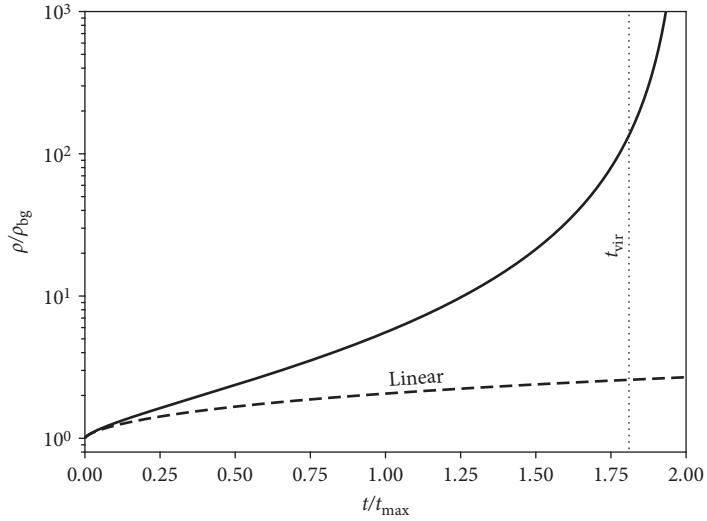


Figure 10.1 The evolution of the density of a gravitationally collapsing region. The dashed curve is the behavior predicted in the linear regime (valid only for $t \ll t_{max}$).

where a is the scale factor in the uniform density background, which can be found by solving the first Friedmann equation:

$$\frac{da}{dt}\frac{1}{a} = \left(\frac{8\pi G\rho_{crit,0}\, a^{-3}}{3}\right)^{1/2} \tag{10.38}$$

$$\longrightarrow \left(\frac{3}{8\pi G\rho_{crit,0}}\right)^{1/2} \int_0^a \sqrt{a'}\, da' = \int_0^{t_{max}} dt$$

$$\longrightarrow a = \left(\frac{3}{2}\right)^{2/3} t_{max}^{2/3} \left(\frac{8\pi G\rho_{crit,0}}{3}\right)^{1/3}.$$

This calculation shows that an overdense region will stop expanding and begin to collapse when its density exceeds that of the surrounding volume by a factor of approximately $9\pi^2/16 \approx 5.6$. Notice that this is a much greater density contrast than would have been predicted by a naïve extrapolation of linear theory, $1 + \delta_{linear} = 1 + (3/20)(9t_{max}/2\pi G\rho_0)^{2/3} = 1 + (3/20)(6\pi)^{2/3} \approx 2.06$ (see Figure 10.1).

Eqs. 10.32 and 10.33 further suggest that the region in question will collapse into an infinite density singularity ($a \to 0$) when $\eta = 2\pi$, corresponding to $t = 2t_{max} = 8\pi^2 G\rho_0/3$. In reality, this extreme behavior is prevented by the virialization of the halo, in which the kinetic energy of the collapsing matter is converted into random motion. It is at this time of virialization that the overdensity can be said to have collapsed into a halo, at which point the linear density contrast is given by $\delta_{linear} \approx (3/20)(12\pi)^{2/3} \approx 1.686$ [82].

To estimate the density of a halo at the time of its collapse and virialization, we cannot rely on linear theory. Instead, we can consider the kinetic and potential energies of this system before and after collapse. At the start of collapse, $t = t_{max}$, the total energy of the system is given by the potential energy, E_p (the kinetic energy is zero when the scale factor is equal to

its maximum value), and is equal to

$$E_{\rm tot} = E_p = -\int_0^{a_{\rm max}} \frac{GM(<r)}{r}\frac{dM}{dr}dr \tag{10.39}$$
$$= -\frac{16\pi^2\rho^2 G}{3}\int_0^{a_{\rm max}} r^4 dr$$
$$= -\frac{9M^2}{40\pi\rho_0},$$

where $M(<r) = 4\pi\rho r^3/3$, $dM/dr = 4\pi r^2\rho$, and $M \equiv M(<a_{\rm max})$ is the total mass contained in the region. In the last step, I have made use of $a_{\rm max} = 8\pi G\rho_0/3$.

After collapse and virialization, we can use the virial theorem to equate the total kinetic energy to $-1/2$ times the potential energy:

$$E_{\rm tot} - E_p = -\frac{1}{2}E_p \tag{10.40}$$
$$-\frac{9M^2}{40\pi\rho_0} + \frac{3GM^2}{5a_{\rm vir}} = \frac{3GM^2}{10a_{\rm vir}},$$

where I have used the fact that the gravitational potential energy of a spherical body of uniform density is given by $-3GM^2/5r$. From this, it follows that $a_{\rm vir} = 4\pi G\rho_0/3 = a_{\rm max}/2$. In other words, the overdense region virializes when its radius has decreased to half of its maximum value. This occurs when $\eta = 3\pi/2$, corresponding to $t_{\rm vir} = 4\pi G\rho_0\,(3\pi/2+1)/3 = t_{\rm max}\,(3/2+1/\pi) \approx 1.81\,t_{\rm max}$. Between $t_{\rm max}$ and $t_{\rm vir}$, the scale factor of the surrounding background increases by a factor of $(t_{\rm vir}/t_{\rm max})^{2/3}$ (assuming matter domination), reducing the density by a factor of $[(t_{\rm vir}/t_{\rm max})^{2/3}]^3 \approx 3.28$. Combining these factors, the density of the collapsing region at the time of virialization is greater than the background density by a factor of

$$\left.\frac{\rho}{\rho_{\rm bg}}\right|_{t_{\rm vir}} \sim \left.\frac{\rho}{\rho_{\rm bg}}\right|_{t_{\rm max}} \times \left(\frac{a_{\rm max}}{a_{\rm vir}}\right)^3 \times \left(\frac{t_{\rm vir}}{t_{\rm max}}\right)^2 \tag{10.41}$$
$$= \frac{9\pi^2}{16} \times 2^3 \times 1.81^2 \approx 145.$$

Thus density perturbations form gravitationally bound halos when their density exceeds the average density of the universe by a factor of roughly $\sim 100 - 200$. In practice, virialization is generally taken to occur when $\rho = 180\,\rho_{\rm bg}$. Despite the approximations used in this calculation (sphericity, matter domination), the essence of this result has been confirmed by numerical simulations.

We can use this result to estimate the redshift at which a clump of a given mass will undergo virialization. We start by rewriting Eq. 10.41 as the following condition for virialization:

$$\rho \gtrsim 180\rho_{\rm crit}\,\Omega_M\,(1+z_{\rm vir})^3. \tag{10.42}$$

Next, we can use the virial theorem to estimate the density of the virialized region in terms of its velocity dispersion, σ,

$$\frac{1}{2}M\sigma^2 \sim \frac{GM^2}{2R} \tag{10.43}$$
$$\longrightarrow \rho \sim \frac{3\sigma^6}{4\pi G^3 M^2}.$$

When these expressions are combined, they yield

$$1 + z_{\rm vir} \lesssim \left(\frac{2}{180}\right)^{1/3} \frac{\sigma^2}{G^{2/3}\, M^{2/3}\, \Omega_M^{1/3}\, H_0^{2/3}} \tag{10.44}$$
$$\lesssim 3.5 \times \left(\frac{\sigma}{220\,{\rm km/s}}\right)^2 \left(\frac{10^{12}\, M_\odot}{M}\right)^{2/3}.$$

While this result should be viewed as only a rough approximation, it does illustrate some important points. First, the largest structures were virialized most recently, with Milky Way-like galaxies ($M \sim 10^{12}\, M_\odot$, $\sigma \sim 220\,{\rm km/s}$) typically forming around or after $z \sim 2$–3, and with massive galaxy clusters ($M \sim 10^{14} - 10^{15}\, M_\odot$, $\sigma \sim$ 500–1,000 km/s) forming closer to the current era, $z \lesssim 2$. Second, massive galaxy clusters will have virialized by the current epoch only if they have large velocity dispersions. If we assume that the gas within a virialized structure reaches hydrostatic equilibrium, and roughly relating $3T/2 \sim m_p\sigma^2/2$, we find that typical clusters should contain gas at a temperature of $T \sim 1 - 10\,{\rm keV}$, consistent with observations.

We are now in a position to estimate the distribution of gravitationally bound clumps at a given redshift, using what is known as the Press-Schechter formalism [83]. To this end, consider the amplitude of a density perturbation averaged over a spherical region of radius R:

$$\bar{\delta}(\vec{x}) = \int d^3x'\, W(|\vec{x} - \vec{x}'|)\, \delta(\vec{x}'), \tag{10.45}$$

where $W(|\vec{x} - \vec{x}'|)$ is a top-hat window function, defined as

$$W(r) \equiv \frac{3}{4\pi R^3} \times \begin{cases} 1 & (r < R), \\ 0 & (r > R). \end{cases} \tag{10.46}$$

Window functions of a Gaussian form are also sometimes used in this context.

In the Press-Schechter approach, one treats any region with $\bar{\delta} > 1.686$ to be a collapsed halo, based on the value of $\delta_{\rm linear}$ at which gravitational collapse occurs. Such a halo will contain a mass equal to $M = (4\pi/3)\bar{\rho}_M R^3$. From this, we can calculate the fraction of the total mass that is bound within halos more massive than $M_{\rm min}$:

$$f(M_{\rm min}) = \frac{1}{\bar{\rho}_M} \int_{M_{\rm min}}^{\infty} dM M \frac{dn}{dM}, \tag{10.47}$$

where dn/dM is the differential number density of gravitationally bound halos. The quantity $f(M_{\rm min})$ can also be viewed as the probability that a randomly chosen region of radius

$R = (3M_{\rm min}/4\pi\,\bar{\rho}_M)^{1/3}$ will have $\bar{\delta} > 1.686$. If the density fluctuations are Gaussian, f can be conveniently written in terms of the complementary error function:

$$f(M_{\rm min}) = \frac{1}{2}\,{\rm erfc}\left(\frac{1.686}{\sigma\sqrt{2}}\right), \tag{10.48}$$

where $\sigma = \langle\bar{\delta}^2\rangle^{1/2}$ is the variance of the density averaged over regions of radius $R = (3M_{\rm min}/4\pi\,\bar{\rho}_M)^{1/3}$. If the spectrum of the density fluctuations extends to arbitrarily small masses, as expected in the purely cold dark matter limit, it can be shown that f approaches 1/2 as $M_{\rm min}$ approaches 0. After all, half of the mass, by definition, is found in regions with $\delta > 0$. In reality, however, much more than half of the total mass ultimately ends up within gravitationally bound structures. For example, mass that starts out in regions of lower than average density can be readily accreted onto high density halos. In light of this, it is traditional to double the number density of halos predicted by the Press-Schechter approach, yielding the following result:

$$\begin{aligned}\frac{dn}{dM} &= \frac{2\bar{\rho}_M}{M}\left|\frac{\partial f(M)}{\partial M}\right| \\ &= \frac{2\bar{\rho}_M}{M}\,\frac{1.686}{\sqrt{2\pi}\sigma^2}\left|\frac{d\sigma}{dM}\right|\,\exp\left[-\frac{1.686^2}{2\sigma^2}\right].\end{aligned} \tag{10.49}$$

Parameterizing the spectrum of primordial density fluctuations as a power-law, we can write the variance as

$$2\sigma^2 = (1.686)^2 \times \left(\frac{M}{M_\star}\right)^{-(2+n_s)/3}, \tag{10.50}$$

where n_s is the spectral index of the primordial density perturbations and $M_\star \sim 10^{15}\,M_\odot$ is the approximate mass scale of the largest halos being virialized in the present epoch. Notice that in the case of a scale-invariant spectrum, $n_s = 1$, this simplifies to $2\sigma^2 = (1.686)^2\,(M_\star/M)$. Combining this expression for the variance with the mass function given in Eq. 10.49, we arrive at the following result:

$$\frac{dn}{dM} = \frac{2\bar{\rho}_M}{M}\,\frac{(2+n_s)}{6\sqrt{\pi}M}\left(\frac{M}{M_\star}\right)^{(2+n_s)/6}\,\exp\left[-\left(\frac{M}{M_\star}\right)^{(2+n_s)/3}\right]. \tag{10.51}$$

This is the famous Press-Schechter mass function. Despite the various approximations and assumptions that go into its derivation, this result is in reasonably good agreement with those obtained using numerical simulations [84].

10.6 Hot, warm, and cold dark matter

In this section, we'll explore how the transfer function, and thus the matter power spectrum, is impacted by the choice of cosmological parameters, and specifically by whether the dark matter is cold, warm, or hot.

Central to the impact of dark matter on the shape of the transfer function is the question of when the perturbations on a given scale entered the horizon (i.e., when the proper

wavelength of a given mode became smaller than the size of the cosmological horizon). From the definition of the comoving horizon in Sec. 2.8, we can see how this quantity evolves during the radiation- and matter-dominated eras:

$$d_h = \int_0^a \frac{da'}{H(a')\,a'^2} \propto \begin{cases} a & \text{(radiation domination)}, \\ a^{1/2} & \text{(matter domination)}. \end{cases} \tag{10.52}$$

Thus, as the universe expanded, density perturbations of increasingly large comoving wavelength entered the horizon.

For the purposes of this discussion, it will be useful to distinguish modes that entered the horizon before the time of matter-radiation equality from those that did so after. With this in mind, consider the mode whose proper wavelength was equal to the Hubble length at the time of equality, $k = k_{\rm Eq}$, where

$$\frac{2\pi\, a_{\rm Eq}}{k_{\rm Eq}} = H_{\rm Eq}^{-1}. \tag{10.53}$$

Solving for $k_{\rm Eq}$, we find

$$\begin{aligned} k_{\rm Eq} &= 2\pi\,(aH)_{\rm Eq} \\ &= \frac{2\sqrt{2}\pi\, H_0\, \Omega_{M,0}^{1/2}}{a_{\rm Eq}^{1/2}} \\ &\approx 0.068\ {\rm Mpc}^{-1} \times \left(\frac{h}{0.7}\right) \left(\frac{\Omega_{M,0}}{0.315}\right)^{1/2}. \end{aligned} \tag{10.54}$$

Notice that the factor of $\sqrt{2}$ in the second line comes from the fact that the energy density at the time of equality was equal to twice the matter density, and thus $H_{\rm Eq}^2 = H_0^2 \times (2\,\Omega_{M,0}/a_{\rm Eq}^3)$.

On sufficiently large scales ($k \lesssim k_{\rm Eq}$), there exist modes that are larger than the Jeans length. These grow as $\delta \propto a^2 \propto t$ during radiation domination and as $\delta \propto a \propto t^{2/3}$ during matter domination (see Eq. 10.24), leading to the following transfer function:

$$T_k \approx \left(\frac{a_{\rm Eq}}{a_i}\right)^2 \frac{a}{a_{\rm Eq}}, \qquad k \lesssim k_{\rm Eq}, \tag{10.55}$$

where a_i is the scale factor at the initial conditions, associated with the primordial power spectrum. Notice that each mode grows by the same amount, regardless of the value of k (so long as $k \lesssim k_{\rm Eq}$).

For modes with $k \gtrsim k_{\rm Eq}$, the size of the horizon becomes larger than the corresponding wavelength before the time of matter-radiation equality. Prior to entering the horizon, the density perturbations associated with these modes grow proportionally to a^2, as described in Eq. 10.24. After entering the horizon, however, they grow only logarithmically (see Eq. 10.20) until the matter-dominated era begins. Treating the slow logarithmic growth as

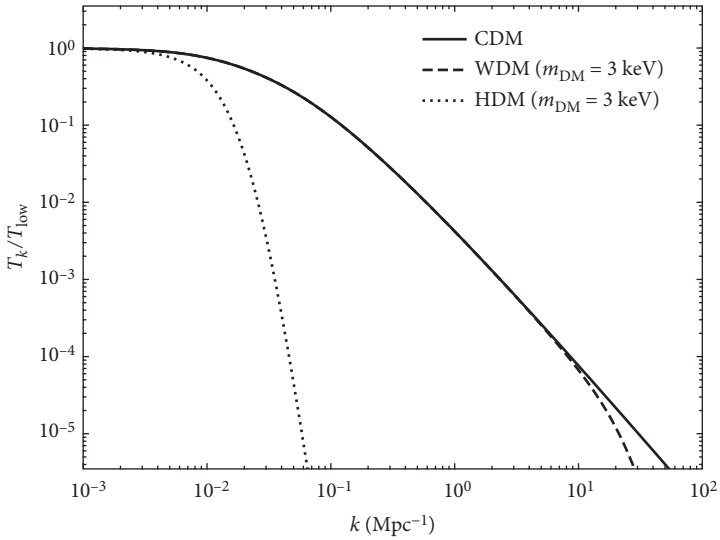


Figure 10.2 The transfer functions for cold, warm, or hot dark matter. In the warm and hot cases, the results shown correspond to a thermal relic with a mass of 3 keV or 3 eV.

approximately constant, the transfer function in this case is given by

$$T_k \approx \left(\frac{a_{\rm enter}}{a_i}\right)^2 \frac{a}{a_{\rm Eq}} \qquad (k \gtrsim k_{\rm Eq}) \tag{10.56}$$
$$= \left(\frac{a_{\rm enter}}{a_{\rm Eq}}\right)^2 \left[\left(\frac{a_{\rm Eq}}{a_i}\right)^2 \frac{a}{a_{\rm Eq}}\right],$$

where $a_{\rm enter}$ is the scale factor at the time that the mode entered the horizon. A mode of wavenumber k is related to the scale factor and Hubble rate at which it enters the horizon by $k = 2\pi(aH)_{\rm enter}$. Using the fact that $H = 1/2t \propto a^{-2}$ during radiation domination, this implies that a given mode enters the horizon when $a_{\rm enter} \propto k^{-1}$. This allows us to substitute $(a_{\rm enter}/a_{\rm Eq})^2 = (k_{\rm Eq}/k)^2$, and thus to write the transfer function as follows:

$$T_k \approx \left(\frac{k_{\rm Eq}}{k}\right)^2 \left[\left(\frac{a_{\rm Eq}}{a_i}\right)^2 \frac{a}{a_{\rm Eq}}\right] \qquad (k \gtrsim k_{\rm Eq}). \tag{10.57}$$

Factors such as free streaming and Silk damping complicate the precise form of the transfer function. Using the fitting formula provided in Ref. [85] to account for these effects, I have plotted the transfer function for the case of cold dark matter in Figure 10.2. As predicted in the discussion above, this transfer function is approximately flat for $k \ll k_{\rm Eq}$, and falls as $T_k \propto k^{-2}$ for all $k \gg k_{\rm Eq}$.

If the dark matter is not perfectly cold, the formation of structure will be suppressed on small scales, corresponding to large k. To understand why this is the case, consider the expression for the Jeans length associated with dark matter, as given in Eq. 10.15. In deriving that equation, we assumed that the dark matter was nonrelativistic at the time of its kinetic decoupling, allowing it to cool as $T \propto (1+z)^2$ from that point onward. If it had instead been relativistic, its temperature would have evolved as $T \propto (1+z)$ until $T_{\rm DM} \sim m_{\rm DM}$. As a result,

the dark matter will be hotter by a factor of $\sim T_{\rm kd}/m_{\rm DM}$. The Jeans length as evaluated in the radiation-dominated era is in this case given by:

$$\begin{aligned}\lambda_J &= \left(\frac{\pi\, T_{\rm DM}}{G\bar{\rho}\, m_{\rm DM}}\right)^{1/2} \\ &\approx \left(\frac{\pi\, T_{\rm kd}}{G\,\Omega_{\rm rad}\,\rho_{\rm crit}(1+z)^4 m_{\rm DM}}\,\frac{(1+z_{\rm nr})}{(1+z_{\rm kd})}\,\frac{(1+z)^2}{(1+z_{\rm nr})^2}\right)^{1/2} \\ &\approx \left(\frac{\pi\, T_{\rm CMB}^4\, g_{\star,S}^{4/3}(T_{\rm CMB})}{G\,\Omega_{\rm rad}\,\rho_{\rm crit} m_{\rm DM}^2\, T^2\, g_{\star,S}^{2/3}(T)\, g_{\star,S}^{1/3}(T_{\rm kd})\, g_{\star,S}^{1/3}(m_{\rm DM})}\right)^{1/2} \\ &\approx 4\times 10^{14}\ {\rm cm}\times\left(\frac{1\ {\rm MeV}}{T}\right)\left(\frac{1\ {\rm keV}}{m_{\rm DM}}\right)\left(\frac{g_{\star,S}^4(T_{\rm CMB})}{g_{\star,S}^2(T)\, g_{\star,S}(T_{\rm kd})\, g_{\star,S}(m_{\rm DM})}\right)^{1/6},\end{aligned} \tag{10.58}$$

where $z_{\rm nr} \approx (m_{\rm DM}/T_{\rm CMB})\,[g_{\star,S}(T_{\rm kd})/g_{\star,S}(T_{\rm CMB})]^{1/3}$ is the redshift at which the dark matter becomes nonrelativistic. A region with a radius equal to the Jeans length encloses the following mass of dark matter:

$$\begin{aligned}M_J &= \frac{4\pi}{3}\lambda_J^3\,\rho_{\rm DM} \\ &= \frac{4\pi^{5/2}\Omega_{\rm DM} T_{\rm CMB}^3 g_{\star,S}(T_{\rm CMB})}{3\rho_{\rm crit}^{1/2} G^{3/2}\Omega_{\rm rad}^{3/2} m_{\rm DM}^3 g_{\star,S}^{1/2}(T_{\rm kd})\, g_{\star,S}^{1/2}(m_{\rm DM})} \\ &\approx 4\times 10^9\, M_\odot\times\left(\frac{1\ {\rm keV}}{m_{\rm DM}}\right)^3\left(\frac{10}{g_{\star,S}(T_{\rm kd})}\right)^{1/2}\left(\frac{3.9}{g_{\star,S}(m_{\rm DM})}\right)^{1/2}.\end{aligned} \tag{10.59}$$

From this, we see that a thermal relic with a mass near the keV scale will be too hot to form gravitationally bound structures smaller than a few billion solar masses. Physically speaking, this is because the dark matter particles can free stream out of the gravitational potentials, escaping from the density perturbations and thereby suppressing the growth of those structures. Notice that the characteristic mass scale below which this suppression occurs is proportional to $1/m_{\rm DM}^3$, and has only a mild dependance on the temperature of kinetic decoupling. This is very different from the result found in the case of a thermal relic that freezes out while nonrelativistic (see Eq. 10.16).

For hot dark matter candidates, including Standard Model neutrinos and other thermal relics with sub-keV masses, the growth of structure is highly suppressed on the scale of galaxies, and perhaps even galaxy clusters [86, 87]. Numerical simulations show that hot dark matter particles tend to collapse and form very large structures first, and only later go on to form smaller, galaxy-sized structures through the fragmentation of larger halos. This sequence is opposite to that predicted in the case of cold dark matter, in which the smallest halos form first and then undergo a succession of mergers to produce increasingly massive structures. By the mid-1980s, measurements of our universe's galaxy distribution [88] had made it clear that most of the dark matter is not hot [89].

The quantitative effects of warm dark matter on the transfer function can be conveniently parameterized as [90, 91]

$$T_{k,{\rm WDM}} \approx T_{k,{\rm CDM}}\left[1+(\alpha k)^{2.24}\right]^{-4.46}, \tag{10.60}$$

where $T_{k,\rm CDM}$ is the transfer function found in the cold dark matter case, and

$$\alpha \approx 0.07\,{\rm Mpc} \times \left(\frac{m_{\rm DM}}{1\,{\rm keV}}\right)^{-1.11}. \tag{10.61}$$

In Figure 10.2, the transfer function for cold dark matter is compared to that predicted for a thermal relic with a mass of either 3 keV or 3 eV, representing examples of warm and hot dark matter, respectively. In the warm dark matter case, structures smaller than $k \sim 8\,{\rm Mpc}^{-1}$ are suppressed, corresponding to masses of $M_{\rm min} \sim (4\pi/3)\rho_{\rm crit}\Omega_M\, k^{-3} \sim 3 \times 10^8\, M_\odot$, consistent with the Jeans mass calculated in Eq. 10.59

Strictly speaking, we already know that the dark matter is not purely cold, warm, or hot. Although most of the dark matter must be cold or perhaps warm, a subdominant portion of our universe's energy density consists of neutrinos, which have acted as hot dark matter throughout the course of cosmic history.

10.7 Baryon acoustic oscillations

In the previous section, we derived the basic features of the matter power spectrum that would arise in a universe containing only dark matter. In such a universe, the primordial density fluctuations are enhanced over time uniquely through the effects of gravity. In our universe, however, there is a significant abundance of baryons, which experience forces other than gravity. In particular, on relatively small scales, it is the electromagnetic force that enables baryonic matter to cool and lose angular momentum through the emission of photons, leading to the formation of galaxies, stars, and planets. On much larger scales, the electromagnetic force played an important role by keeping the baryons tightly coupled to the photons up until the time of recombination.

Prior to decoupling, the competing forces of gravity and pressure gave rise to acoustic oscillations that propagated through the baryon-photon fluid, causing regions to undergo alternating phases of contraction and expansion. In Sec. 7.5.3, we found that the largest regions that were in a state of maximum compression at the time of last scattering were $d_s \approx 0.126\,{\rm Mpc}$ in extent, set by the scale of the sound horizon. Since the time of last scattering, these regions have grown as our universe expanded to a current size of $r_s \approx 0.126\,{\rm Mpc} \times (1+z_{\rm ls}) \approx 140\,{\rm Mpc}$. A region of this size contains an amount of mass given by

$$\begin{aligned} M_s &\approx \frac{4\pi}{3}[d_s(1+z_{\rm ls})]^3\,\Omega_M\,\rho_{\rm crit} \\ &\approx 4.5 \times 10^{17}\,M_\odot. \end{aligned} \tag{10.62}$$

Much like the angular power spectrum of the CMB, the acoustic waves that propagated through the baryon-photon fluid prior to decoupling have left their mark on the distribution of matter in our present universe. By studying the matter power spectrum on very large scales, we can hope to measure the impact of these baryon acoustic oscillations.

Baryons impact the transfer function in two important ways. First, they suppress power on scales smaller than the sound horizon. Second, they induce oscillatory features in the matter power spectrum that behave as $\sin(k\,r_s)$, leading to a sequence of peaks (at $k \approx 5\pi/2r_s$, $9\pi/2r_s$, $13\pi/2r_s, \ldots$) and troughs (at $k \approx 3\pi/2r_s$, $7\pi/2r_s$, $11\pi/2r_s, \ldots$). Both of these effects can be seen in Figure 10.3, where I show the transfer function as calculated

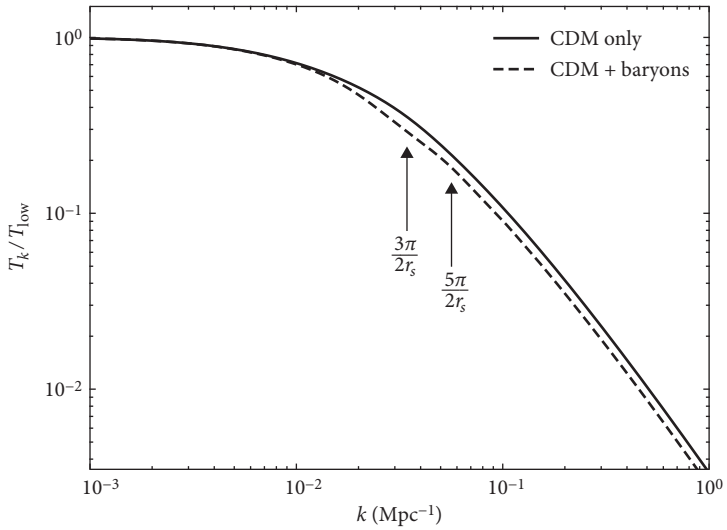


Figure 10.3 The transfer function for cold dark matter with and without including the effects of baryons. I have labeled the wavenumbers at which the first trough ($k = 3\pi/2r_s$) and the first peak ($k = 5\pi/2r_s$) appear in the spectrum of the baryon acoustic oscillations.

for a universe containing only cold dark matter, and including the effects of baryons (as parameterized in Ref. [92]). Note that because most of the matter in our universe is dark matter, the effects of the baryons on the transfer function are rather modest. In a universe with $\Omega_M \sim \Omega_b$, the baryon acoustic oscillations would be much more prominent in the matter power spectrum. Among other implications, this poses a serious challenge for MOND and other modified gravity scenarios that do not include dark matter [93].

Much as cepheids and type 1a supernovae are useful to cosmologists as standard candles, the length scale of the sound horizon can serve as a standard ruler. By this I mean that by measuring the scale at which baryon acoustic oscillations appear in the matter power spectrum as a function of redshift, we can determine how our universe has expanded since the time of recombination. As a consequence, baryon acoustic oscillations provide us with a powerful probe of cosmological parameters, especially those associated with dark energy (see Chapter 14).

The effects of baryon acoustic oscillations were first detected in 2005 by the Sloan Digital Sky Survey (SDSS) [94] (and shortly thereafter by the 2dF Galaxy Redshift Survey [95]), which studied a sample of 46,748 luminous red galaxies at redshifts out to $z = 0.47$. The 2-point correlation function of these galaxies showed a clear bump at the scale of the sound horizon. More recently, several other groups (including the 6dF [96], WiggleZ [97], and BOSS [98] surveys) have detected and studied the effects of baryon acoustic oscillations on the matter power spectrum.

In Figure 10.4, I show the matter power spectrum as calculated in linear theory, using the transfer function from Figure 10.3 (for $\Omega_{\rm DM} = 0.265$, $\Omega_b = 0.049$, and $h = 0.7$) and a primordial spectrum of density perturbations with an index of $n_s = 0.965$, normalized using $\sigma_8 = 0.81$. If you look closely, you should be able to see the wiggles associated with the baryon acoustic oscillations in this curve. The most significant feature of the matter power spectrum is, of course, the break at approximately $k \sim 2 \times 10^{-2}$ [h Mpc^{-1}]. To the left of this break, the modes entered the horizon only after the time of matter-radiation equality, while to the right they entered while our universe was still dominated by radiation.

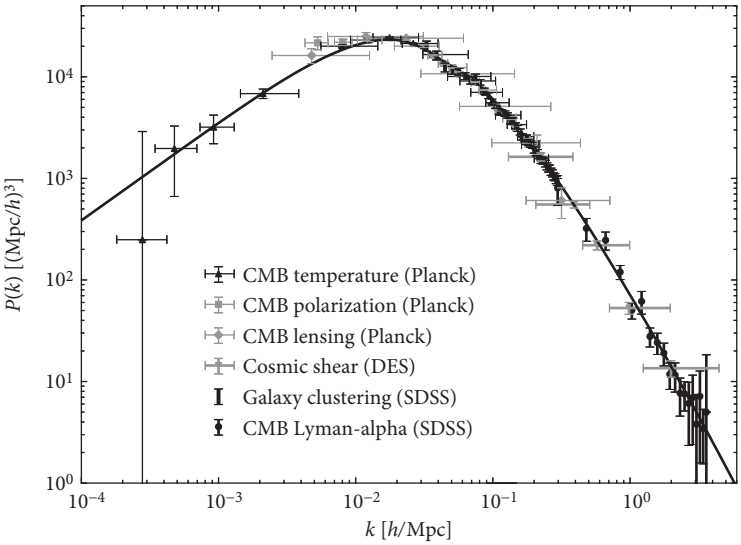


Figure 10.4 The linear-theory matter power spectrum evaluated at $z = 0$ (solid curve). The error bars shown represent the spectrum as inferred from the measurements of Planck, the Dark Energy Survey (DES), and the Sloan Digital Sky Survey (SDSS), as presented in Refs. [99, 100].

The prediction for the linear-theory matter power spectrum is compared in this figure to the values inferred from various cosmological measurements (as compiled in Refs. [99, 100]). On the largest scales, $k \sim 0.0003 - 0.2$ [h/Mpc], the CMB's temperature and polarization power spectra provide a determination that is in good agreement with the predicted shape and amplitude of the matter power spectrum. The gravitational lensing of the CMB by large-scale structure present at $z \sim 0.5 - 10$ also provides a measurement that is in good agreement with these predictions. On intermediate scales, measurements of the matter power spectrum as inferred from galaxy clustering are shown, as reported by the Sloan Digital Sky Survey (SDSS). Also shown are determinations based on measurements of the weak gravitational lensing of distant galaxies by large-scale structure as observed by the Dark Energy Survey (DES), an effect known as "cosmic shear" [101]. On even smaller scales, $k \sim 0.3 - 3.7$ [h/Mpc], measurements derived from the observed intensities of neutral hydrogen absorption lines are shown (as measured by the extended baryon oscillation spectroscopic survey of SDSS). This so-called Lyman-alpha forest provides a particularly strong constraint on warm dark matter, currently limiting $m_{\rm DM} \gtrsim 5\,{\rm keV}$ for the case of a thermal relic [102, 103]. Similarly stringent constraints have also been derived from the observed number of Milky Way satellite galaxies, which is predicted to be suppressed if the dark matter is warm [104]. Note that the quantities $P(k)$ and k have each been scaled in this figure by the Hubble parameter, h, as is often done in presentations of the matter power spectrum.[2]

10.8 Adiabatic and isocurvature perturbations

So far in this chapter, we have taken the primordial density perturbations to be universal across all forms of matter and energy. In other words, we have adopted initial conditions

[2] I would like to thank Marius Millea and Solene Chabanier for providing this compilation of data.

in which the distributions of photons, dark matter, and baryons each trace one another. These kinds of fluctuations are known as adiabatic or curvature perturbations, since they reflect local variations in the total energy density, and thus in the local spatial curvature. It is possible, however, that the primordial density perturbations were not purely adiabatic, but instead featured spatial variations in their equation of state. Such isocurvature perturbations are associated with variations in the *proportions* of the various forms of energy that were present at different locations throughout space.

Adiabatic and isocurvature perturbations evolve very differently, and thus give rise to very different transfer functions. Whereas gravity causes adiabatic perturbations to grow, isocurvature perturbations do not correspond to any variations in the total energy density, and are thus unaffected by gravity. Instead, isocurvature perturbations feature variations in pressure that, on subhorizon scales, can work to redistribute various forms of matter and energy, potentially transforming isocurvature into adiabatic perturbations over time.

Empirically speaking, the angular power spectrum of the CMB does not show any signs of isocurvature [8], but is instead consistent with arising from purely adiabatic initial conditions. This allows us to place strong constraints on a variety of primordial scenarios. We will revisit these constraints when discussing axion dark matter in Sec. 17.3.

Additional reading

For a more detailed treatment of cosmological structure formation, I recommend Dodelson's *Modern Cosmology* and Weinberg's *Cosmology*. There is also an accessible discussion of this topic in the Particle Data Group's review of Big Bang cosmology [24]. For a review of numerical simulations of structure formation, see Ref. [105].

Problems

10.1 Calculate the Jeans length for 1) liquid water, 2) the earth's atmosphere, and 3) gas in the interstellar medium. What is the Jeans mass in each case?

10.2 Consider a dark matter candidate with a mass of 100 GeV that underwent kinetic decoupling when the temperature was $T \sim 1\,\text{keV}$. How did the equation of state of this substance evolve over time?

10.3 From $\lambda_J = (\pi T/G\bar{\rho}\, m)^{1/2}$, derive the expression for the Jeans length for dark matter during radiation domination, as given in Eq. 10.15.

10.4 Consider a thermal relic dark matter candidate with a mass of 1 GeV. What is the maximum value of its kinetic decoupling temperature that would still allow for the formation of clumps with $M \sim M_\odot$?

10.5 Show that Eqs. 10.32 and 10.33 can be rewritten as follows in the small t limit:

$$a \approx \frac{2\pi G\rho_0}{3}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\left[1 - \frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right]. \tag{10.63}$$

10.6 In a closed, matter-dominated universe that is in its contracting phase, how does δ evolve with time? Calculate this for scales larger than and smaller than the Jeans length. Assume that the degree of curvature is small, $\Omega_M \approx 1$.

10.7 Show that Eq. 10.49 follows directly from Eqs. 10.47 and 10.48.

10.8 Show that the Press-Schechter mass function given in Eq. 10.51 follows from Eqs. 10.49 and 10.50.

10.9 How does the size of the horizon evolve during an era dominated by dark energy? Why did I not include this case in Eq. 10.52?

10.10 For a warm dark matter candidate with a mass of 1 keV, how much can the temperature of kinetic decoupling potentially impact the Jeans mass?

10.11 Consider a population of $m_X = 1\,\text{GeV}$ dark matter particles that were produced through a nonthermal mechanism (such as out-of-equilibrium decays) with an average energy of $\langle E_X \rangle \approx 100\,\text{GeV}$ when the temperature of the universe was $T = 1$ GeV. If these dark matter particles never reach kinetic equilibrium with the Standard Model bath, estimate the resulting Jeans mass. What if these particles had instead been produced with $\langle E_X \rangle \approx 10^6\,\text{GeV}$?

11

Neutrino Masses and Oscillations

Unlike charged leptons, neutrinos can be produced with one flavor (ν_e, ν_μ, ν_τ) and later be observed to have another. Such neutrino oscillations are only possible if these particles have nonzero masses, requiring physics beyond the Standard Model. In this short chapter, I do not revisit the role of neutrinos in cosmology, but instead focus on the masses and mixings of these particles, leading to their observed flavor oscillations. While the contents of this chapter might, at first glance, appear to be only tangentially related to the subject of cosmology, neutrino masses and mixings will be essential to our later discussions of leptogenesis (Chapter 12), neutrino astrophysics (Chapter 16), and sterile neutrino dark matter (Chapter 17).

11.1 Neutrino masses

In the Standard Model, the masses of the quarks and charged leptons, along with the W and Z bosons, are generated through the Higgs mechanism. More specifically, fermion masses arise from the Yukawa couplings of the scalar Higgs doublet, ϕ, to the fermion's left-handed and right-handed components. For the charged leptons, for example, this interaction is described by the following terms in the Lagrangian:

$$\mathcal{L} = -y_\ell \, \bar{L}_L \, \phi \, E_R + \text{h.c.}, \tag{11.1}$$

where y_ℓ is the appropriate Yukawa coupling, L_L represents the left-handed $SU(2)$ doublet, E_R is the right-handed singlet, and h.c. stands in for the Hermitian conjugate of the preceding term. After electroweak symmetry breaking, these same terms in the Lagrangian become

$$\mathcal{L} = -y_\ell \frac{v}{\sqrt{2}} \, (\bar{l}_L \, l_R + \bar{l}_R \, l_L), \tag{11.2}$$

where $v \approx 246\,\text{GeV}$ is the vacuum expectation value of the Higgs field. The symbols $l_{L,R}$ represent left- and right-handed leptons, while $\bar{l}_{L,R}$ denote their antiparticles. This interaction gives rise to charged lepton masses that are proportional to the corresponding

Yukawa coupling and the Higgs vev:

$$m_\ell = y_\ell \frac{v}{\sqrt{2}}. \tag{11.3}$$

Interactions of this kind generate masses for each of the quarks and charged leptons. In the Standard Model, however, there are no right-handed neutrinos (or left-handed antineutrinos), making it impossible to generate neutrino masses through the Higgs mechanism. In this sense, the empirical fact that the neutrinos are not massless requires the existence of physics beyond the boundaries of the Standard Model.

To generate masses for the Standard Model neutrinos, we must introduce new particles in the form of right-handed neutrinos. In order to respect the gauge symmetries of the Standard Model, however, these right-handed neutrinos must be fundamentally different from their left-handed counterparts. More specifically, they cannot couple to the W or Z bosons, and thus do not experience the weak force. Neutrinos which do not participate in the weak interactions are known as *sterile* neutrinos, in contrast to their left-handed, *active* counterparts.

Once we have introduced sterile neutrinos, we can write down a gauge-invariant interaction similar to that given in Eq. 11.1, which results in the generation of masses for the Standard Model neutrinos:

$$\begin{aligned} \mathcal{L} &= -y_\nu \bar{L}_L \phi N_R + \text{h.c.} \\ &= -m_D (\bar{\nu}_L N_R + \bar{N}_R \nu_L), \end{aligned} \tag{11.4}$$

where N_R represents the sterile neutrino, and m_D is the Dirac mass that is generated. Alternatively, in the presence of sterile neutrinos, we could also generate masses for the Standard Model neutrinos through interactions of the following form:

$$\mathcal{L} = -\frac{1}{2} m_L \bar{\nu}_L \nu_R^C - \frac{1}{2} m_R \bar{N}_L^C N_R + \text{h.c.}, \tag{11.5}$$

where the superscript C represents the charge conjugate of the labeled field. The symbol ν_R^C, for example, denotes a right-handed antineutrino, while N_L^C represents a left-handed, sterile antineutrino.[1] The quantities m_L and m_R are known as Majorana masses. Interactions such as these allow for the creation and destruction of lepton number, and will be important in our discussion of leptogenesis in Sec. 12.5.

Combining the interactions described in Eqs. 11.4 and 11.5, we can write the terms of the Lagrangian that generate neutrino masses as follows:

$$\begin{aligned} \mathcal{L} &= -\frac{1}{2} m_D (\bar{\nu}_L N_R + \bar{N}_L^C \nu_R^C) - \frac{1}{2} m_L \bar{\nu}_L \nu_R^C - \frac{1}{2} m_R \bar{N}_L^C N_R + \text{h.c.} \\ &= -\frac{1}{2} (\bar{\nu}_L, \bar{N}_L^C) \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} \nu_R^C \\ N_R \end{pmatrix} + \text{h.c.} \end{aligned} \tag{11.6}$$

To find the resulting eigenstates, we diagonalize the 2×2 mass matrix in the expression above, obtaining the following:

[1] Be warned that this notation can vary significantly between authors.

$$\nu_{1L} = \cos\theta\ \nu_L - \sin\theta\ N_L^C, \tag{11.7}$$
$$\nu_{1R}^C = \cos\theta\ \nu_R^C - \sin\theta\ N_R$$

and

$$\nu_{2L} = \sin\theta\ \nu_L + \cos\theta\ N_L^C,$$
$$\nu_{2R}^C = \sin\theta\ \nu_R^C + \cos\theta\ N_R,$$

where the mixing angle is given by $\tan 2\theta = 2m_D/(m_R - m_L)$. Expressing these equations in words, we now have two pairs of Majorana fields, each consisting of a left-handed neutrino, ν_{iL}, and a right-handed antineutrino, ν_{iR}^C. Furthermore, each of these Majorana fields can be written as a mixture of the active and sterile states that we started with:

$$\phi_1 = \nu_{1L} \pm \nu_{1R}^C, \tag{11.8}$$
$$\phi_2 = \nu_{2L} \pm \nu_{2R}^C,$$

where the masses of these two particles are given by

$$m_{1,2} = \frac{1}{2}\left| m_L + m_R \pm \sqrt{(m_L - m_R)^2 + 4m_D^2} \right|. \tag{11.9}$$

At this point, it may be illuminating to consider a few specific limiting cases:

1. **The Dirac case**: If no Majorana masses are generated ($m_L = m_R = 0$), the Dirac mass term results in two degenerate states, each with $m_{1,2} = m_D$. Furthermore, the mixing angle is maximal in this case, with $\sin\theta = \cos\theta = 1/\sqrt{2}$. Since both states have the same mass, they combine to form a single Dirac field, $\nu_L + N_R$.
2. **The Majorana case**: In the absence of any Dirac mass term ($m_D = 0$), the two mass eigenstates are simply given by $m_{1,2} = m_{L,R}$. Furthermore, the mixing angle is zero in this case, allowing us to write the two independent Majorana fields as $\phi_1 = \nu_L + \nu_R^C$ and $\phi_2 = N_L^C + N_R$.
3. **The pseudo-Dirac case**: If the Dirac mass is much larger than the Majorana masses ($m_D \gg m_L, m_R$), the two mass eigenstates will be nearly degenerate, with $m_{1,2} \approx m_D \pm (m_L + m_R)$. As in the purely Dirac case, the mixing angle will be nearly maximal, leading to Majorana fields that collectively constitute a single pseudo-Dirac state.
4. **The seesaw case**: If the right-handed Majorana mass is much larger than the Dirac mass and no left-handed Majorana mass is generated ($m_R \gg m_D$, $m_L = 0$), there will be one very light and one very heavy mass eigenstate, $m_1 \approx m_D^2/m_R$ and $m_2 \approx m_R$. Mixing in this case is highly suppressed, $\theta \approx m_D/m_R$, and thus the heavy state is almost purely sterile, while the light state is almost purely active. This kind of scenario can naturally arise within the context of grand unified theories, and provides an attractive explanation for why the active neutrinos are so light [106, 107].

11.2 Neutrino mixing and oscillations

If one came across a charged lepton with a mass of 511 keV, there would be no doubt that this particle is an electron. In contrast, a neutrino with a well-defined mass does not possess a unique flavor, but instead exists in a state that consists of a mixture of flavors. To put this another way, the mass eigenstates of the neutrinos are not the same as their flavor eigenstates. This is closely analogous to how the CKM matrix (see Eq. 3.1) relates the mass eigenstates of the quarks to their flavor eigenstates.

We can write each neutrino mass eigenstate as a sum of their flavor eigenstates,

$$|\nu_i\rangle = \sum_\alpha U_{\alpha i} |\nu_\alpha\rangle, \tag{11.10}$$

or the flavor eigenstates as a sum of the mass eigenstates,

$$|\nu_\alpha\rangle = \sum_i U^\star_{\alpha i} |\nu_i\rangle, \tag{11.11}$$

where in each case $\alpha = e, \mu, \tau$ labels the flavor eigenstates and $i = 1, 2, 3$ labels the mass eigenstates. For antineutrinos, we instead relate the mass and flavor eigenstates as follows:

$$|\bar{\nu}_i\rangle = \sum_\alpha U^\star_{\alpha i} |\bar{\nu}_\alpha\rangle, \qquad |\bar{\nu}_\alpha\rangle = \sum_i U_{\alpha i} |\bar{\nu}_i\rangle. \tag{11.12}$$

The neutrino mixing matrix, $U_{\alpha i}$, is unitary (i.e., $U^\dagger U = 1$), and has the following properties:

$$\sum_i U_{\alpha i} U^*_{\beta i} = \delta_{\alpha\beta}, \qquad \sum_\alpha U_{\alpha i} U^*_{\alpha j} = \delta_{ij}. \tag{11.13}$$

In the case of three Majorana neutrinos, this mixing matrix can be parameterized in terms three mixing angles ($\theta_{12}, \theta_{23}, \theta_{13}$) and three phases ($\delta_{\rm CP}, \eta_1, \eta_2$). In this parameterization, we can write U as the following product of matrices:

$$\begin{aligned} U &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & c_{23} & s_{23} \\ 0 & -s_{23} & c_{23} \end{pmatrix} \begin{pmatrix} c_{13} & 0 & s_{13} e^{-i\delta_{\rm CP}} \\ 0 & 1 & 0 \\ -s_{13} e^{i\delta_{\rm CP}} & 0 & c_{13} \end{pmatrix} \begin{pmatrix} c_{12} & s_{12} & 0 \\ -s_{12} & c_{12} & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} e^{i\eta_1} & 0 & 0 \\ 0 & e^{i\eta_2} & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} c_{12} c_{13} e^{i\eta_1} & s_{12} c_{13} e^{i\eta_2} & s_{13} e^{-i\delta_{\rm CP}} \\ -s_{12} c_{23} e^{i\eta_1} - c_{12} s_{13} s_{23} e^{i\eta_1} e^{i\delta_{\rm CP}} & c_{12} c_{23} e^{i\eta_2} - s_{12} s_{13} s_{23} e^{i\eta_2} e^{i\delta_{\rm CP}} & c_{13} s_{23} \\ s_{12} s_{23} e^{i\eta_1} - c_{12} s_{13} c_{23} e^{i\eta_1} e^{i\delta_{\rm CP}} & -c_{12} s_{23} e^{i\eta_2} - s_{12} s_{13} c_{23} e^{i\eta_2} e^{i\delta_{\rm CP}} & c_{13} c_{23} \end{pmatrix}, \end{aligned} \tag{11.14}$$

where $c_{ij} \equiv \cos\theta_{ij}$ and $s_{ij} \equiv \sin\theta_{ij}$.

If the three neutrinos are Dirac, the phases η_1 and η_2 can be absorbed into the definitions of the neutrino states, allowing us to write the mixing matrix as follows:

$$U = \begin{pmatrix} c_{12} c_{13} & s_{12} c_{13} & s_{13} e^{-i\delta_{\rm CP}} \\ -s_{12} c_{23} - c_{12} s_{13} s_{23} e^{i\delta_{\rm CP}} & c_{12} c_{23} - s_{12} s_{13} s_{23} e^{i\delta_{\rm CP}} & c_{13} s_{23} \\ s_{12} s_{23} - c_{12} s_{13} c_{23} e^{i\delta_{\rm CP}} & -c_{12} s_{23} - s_{12} s_{13} c_{23} e^{i\delta_{\rm CP}} & c_{13} c_{23} \end{pmatrix}. \tag{11.15}$$

In this form, U is known as the Pontecorvo-Maki-Nakagawa-Sakata (PMNS) matrix.

The neutrino mass eigenstates are quantum mechanical stationary states, and thus evolve according to the plane wave solution of the Schrödinger equation:

$$\begin{aligned} |\nu_i(x,t)\rangle &= e^{-iE_it}\,|\nu_i(x,0)\rangle \\ &= e^{-iE_it}\,e^{ipx}\,|\nu_i\rangle, \end{aligned} \tag{11.16}$$

where E_i is the energy of the ith mass eigenstate.

Consider a neutrino that is produced at $x=0$ and $t=0$ through a charged current interaction with a charged lepton of a particular flavor α. This neutrino is initially in an eigenstate of flavor α, as described by Eq. 11.11, but will evolve over time into a mixture of flavors according to

$$\begin{aligned} |\nu(x,t)\rangle &= \sum_i U^\star_{\alpha i}\,e^{ipx}e^{-iE_it}\,|\nu_i\rangle \\ &= \sum_i U^\star_{\alpha i}\,e^{ipx}e^{-iE_it}\sum_\beta U_{\beta i}\,|\nu_\beta\rangle. \end{aligned} \tag{11.17}$$

This allows us to write the amplitude for the transition $\alpha \to \beta$ as

$$\begin{aligned} \langle\nu_\beta|\nu(x,t)\rangle &= \sum_i U^\star_{\alpha i}U_{\beta i}\,e^{ipx}e^{-iE_it} \\ &= \sum_i U^\star_{\alpha i}U_{\beta i}\,e^{-im_i^2L/2E}, \end{aligned} \tag{11.18}$$

where $L=x=ct$, and in the last line I have assumed that the neutrino is relativistic, $E_i=\sqrt{m_i^2+p^2}\approx p+m_i^2/2p$, allowing me to write:

$$\begin{aligned} ipx-iE_it &\approx ipL-i\left(p+\frac{m_i^2}{2E}\right)L \\ &\approx -\frac{im_i^2L}{2E}. \end{aligned} \tag{11.19}$$

The probability that a neutrino which starts out in a pure state, α, will interact as if it were a neutrino of flavor, β, after propagating a distance, L, is given by

$$\begin{aligned} P_{\nu_\alpha\to\nu_\beta} &= \left|\langle\nu_\beta|\nu_\alpha\rangle\right|^2 \\ &= \left|\sum_i U^\star_{\alpha i}U_{\beta i}\,e^{-im_i^2L/2E}\right|^2 \\ &= \sum_i U^\star_{\alpha i}U_{\beta i}\,e^{-im_i^2L/2E}\sum_j U_{\alpha j}U^\star_{\beta j}\,e^{im_j^2L/2E} \\ &= \sum_i |U_{\alpha i}|^2|U_{\beta i}|^2+\sum_{i\neq j}U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,e^{-i(m_i^2-m_j^2)L/2E}. \end{aligned} \tag{11.20}$$

Using Euler's formula and the double angle formula, we can write

$$e^{-i(m_i^2-m_j^2)L/2E} = \cos\left(-\frac{(m_i^2-m_j^2)L}{2E}\right) + i\sin\left(-\frac{(m_i^2-m_j^2)L}{2E}\right) \tag{11.21}$$
$$= 1 - 2\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right) - i\sin\left(\frac{(m_i^2-m_j^2)L}{2E}\right),$$

allowing us to express the oscillation probability as follows:

$$P_{\nu_\alpha\to\nu_\beta} = \sum_i |U_{\alpha i}|^2|U_{\beta i}|^2 + \sum_{i\neq j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j} - 2\sum_{i\neq j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right)$$
$$- i\sum_{i\neq j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,\sin\left(\frac{(m_i^2-m_j^2)L}{2E}\right) \tag{11.22}$$
$$= \delta_{\alpha\beta} - 4\sum_{i>j}\mathrm{Re}(U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j})\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right)$$
$$+ 2\sum_{i>j}\mathrm{Im}(U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j})\,\sin\left(\frac{(m_i^2-m_j^2)L}{2E}\right).$$

For antineutrinos, this equation is unchanged except that the imaginary term is opposite in sign. Note that in the absence of CP violation ($\delta_{\mathrm{CP}} = 0$), the product $U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}$ is real, and thus the last term in this equation vanishes.

While the amplitude for neutrino oscillations is determined by the sum of the products of entries in the mixing matrix, the frequency at which these oscillations occur is entirely determined by the mass splittings between the mass eigenstates, through the quantity $\Delta m^2_{ij} \equiv m_i^2 - m_j^2$. In particular, the oscillations undergo a complete cycle, returning the oscillation probability to zero when $\Delta m^2_{ij}L/2E = 2\pi$. This motivates the introduction of a quantity known as the oscillation length:

$$L_0 \equiv \frac{4\pi E}{|\Delta m^2_{ij}|} \approx 992\,\mathrm{m} \times \left(\frac{E}{\mathrm{MeV}}\right)\left(\frac{0.0025\,\mathrm{eV}^2}{|\Delta m^2_{ij}|}\right). \tag{11.23}$$

Once a non-monoenergetic beam of neutrinos has traveled a distance much greater than the oscillation length, $L \gg L_0$, the neutrinos will tend to evolve into a state with flavor probabilities that can be calculated using the time-averaged values of the trigonometric factors, $\langle\sin^2(\Delta m^2 L/4E)\rangle = \langle\sin^2(\Delta m^2 L/2E)\rangle = 0.5$.

The full expression for the oscillation probabilities given in Eq. 11.22 can be a bit cumbersome. For pedagogical purposes, let's consider the simplified case of oscillations between two flavors of neutrinos, with a mixing matrix given by:

$$U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}. \tag{11.24}$$

Applying Eq. 11.22 to this mixing matrix, we find the following oscillation probabilities:

$$\begin{aligned} P_{\nu_\alpha \to \nu_\beta} &= 4\cos^2\theta \ \sin^2\theta \ \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \\ &= \sin^2(2\theta) \ \sin^2\left(\frac{\Delta m^2 L}{4E}\right), \\ P_{\nu_\alpha \to \nu_\alpha} &= 1 - 4\cos^2\theta \ \sin^2\theta \ \sin^2\left(\frac{\Delta m^2 L}{4E}\right) \\ &= 1 - \sin^2(2\theta) \ \sin^2\left(\frac{\Delta m^2 L}{4E}\right), \end{aligned} \tag{11.25}$$

where $\alpha \neq \beta$, and Δm^2 represents the difference between the squared masses of the two neutrinos.

By measuring neutrinos of different energies that have propagated over different length scales, one can measure the various mass splittings, mixing angles, and CP-violating phases of the Standard Model neutrinos. The measured mass splittings are given by

$$\begin{aligned} \Delta m^2_{21} &\approx 7.5 \times 10^{-5}\,\text{eV}^2, \\ \Delta m^2_{32} &\approx \pm 2.5 \times 10^{-3}\,\text{eV}^2, \end{aligned} \tag{11.26}$$

which correspond to oscillations that occur on length scales of $L_0 \sim 30\,\text{km} \times (E/\text{MeV})$ and $L_0 \sim 1\,\text{km} \times (E/\text{MeV})$, respectively. Note that Δm^2_{32} could be either positive or negative. In the positive case, we have $m_1 < m_2 < m_3$, which is known as the normal hierarchy. In the negative case, we instead have the inverted mass hierarchy, in which $m_3 < m_1 < m_2$.

The measured mixing angles of the Standard Model neutrinos are given by

$$\begin{aligned} \sin^2\theta_{23} &\approx 0.52 - 0.56, \\ \sin^2\theta_{12} &\approx 0.31, \\ \sin^2\theta_{13} &\approx 0.022. \end{aligned} \tag{11.27}$$

Oscillations associated with these mass splittings and mixing angles have been observed on a variety of length scales, including among neutrinos produced in cosmic-ray interactions in the atmosphere [108], and through nuclear fusion in the sun [109, 110].

11.3 Neutrino oscillations in the presence of matter

In the previous section, we explored how neutrinos oscillate as they propagate through empty space. In the presence of a dense medium, such as that found within a star or in the early universe, neutrino oscillations can proceed quite differently. In particular, the interactions of neutrinos with the particles in their environment can alter their effective mixing angles and mass splittings, significantly impacting the rates at which they undergo flavor oscillations [111, 112, 113].

The effects of matter on neutrino oscillations arise from the fact that different flavors of neutrinos experience different interactions with matter. Electron neutrinos, for example, can interact with electrons through charged current diagrams that do not exist for muon or tau

neutrinos. This difference leads to a potential given by $V = \sqrt{2} G_F n_e$ for electron neutrinos and $V = -\sqrt{2} G_F n_e$ for electron antineutrinos, where n_e is the number density of electrons in the environment.

Consider, for simplicity, the oscillations of electron neutrinos with neutrinos of one other flavor. In the presence of matter, the effective mass splitting and mixing angle, Δm^2_m and θ_m, can be written in terms of these parameters in vacuum, the neutrino's energy, and the potential:

$$\Delta m^2_m = \Delta m^2 \sqrt{\sin^2 2\theta + (\cos 2\theta - 2EV/\Delta m^2)^2} \tag{11.28}$$

and

$$\sin^2 2\theta_m = \frac{\sin^2 2\theta}{\sin^2 2\theta + (\cos 2\theta - 2EV/\Delta m^2)^2}, \tag{11.29}$$

leading to an oscillation probability that is given by

$$P_{\nu_e \to \nu_\mu} = \sin^2 2\theta_m \, \sin^2 \left(\frac{\Delta m^2_m L}{4E} \right). \tag{11.30}$$

On short baselines, $L \ll 4E/\Delta m^2_m$, we can use the small angle approximation to express the oscillation probability as follows:

$$\begin{aligned} P_{\nu_e \to \nu_\mu} &\approx \sin^2 2\theta_m \left(\frac{\Delta m^2_m L}{4E} \right)^2 \\ &= \sin^2 2\theta \left(\frac{\Delta m^2 L}{4E} \right)^2, \end{aligned} \tag{11.31}$$

which is equal to the probability obtained in vacuum. Matter effects thus have a significant impact only on relatively long baselines, $L \gtrsim 4E/\Delta m^2_m$.

Notice that matter effects are particularly important when $E \sim \Delta m^2 \cos 2\theta / 2V$. Near this resonance, $\sin 2\theta_m \sim 1$, regardless of the value of the mixing angle in vacuum. Furthermore, the effective mass splitting is suppressed for neutrinos near this resonance, $\Delta m^2_m \sim \Delta m^2 \, \sin 2\theta$. Neutrinos with this resonant energy thus oscillate more rapidly and with a greater amplitude.

In the case of neutrinos propagating through the sun, the relevant potential is generated almost entirely through the charged current scattering of neutrinos with electrons. As we will discuss in Sec. 17.2, matter effects may also have played an important role in the production of sterile neutrinos in the early universe.

Additional reading

I've long been fond of Zuber's *Neutrino Physics*, as well as Giunti and Kim's *Neutrino Physics and Astrophysics*. For a deeper dive into the impact of neutrinos on the CMB and large-scale structure, I recommend *Neutrino Cosmology* by Lesgourgues, Mangano, Miele, and Pastor. De Gouvea's lectures on neutrino physics from the TASI summer school [114] is another great resource on this topic. The Particle Data Group's *Review of Particle Physics* [24] also includes a useful review of this subject.

Problems

11.1 A Dirac mass term is sometimes written as $\mathcal{L} = -m_D \bar{\psi}\psi$, where $\psi = \psi_L + \psi_R$ and $\bar{\psi} = \bar{\psi}_L + \bar{\psi}_R$. Show that this is equivalent to the expression given in Eq. 11.2.

11.2 By performing a Taylor expansion of Eq. 11.9 in the seesaw case ($m_R \gg m_D$, $m_L = 0$), show that one obtains mass eigenstates that are approximately given by $m_1 \approx m_D^2/m_R$ and $m_2 \approx m_R$.

11.3 Show that the final result in Eq. 11.22 follows from the first equality in that expression.

11.4 In many instances, the phenomenology of neutrino oscillations is dominated by a single mass splitting. In such situations, the expressions for the oscillation probabilities simplify considerably. For example, in terms of $\Delta m^2_{\text{atm}} \approx \Delta m^2_{31} \approx \Delta m^2_{32} \gg \Delta m^2_{21}$, show that the oscillation probabilities can be written as:

$$P_{\nu_e \to \nu_\tau} \approx \sin^2(2\theta_{13}) \cos^2(\theta_{23}) \ \sin^2\left(\frac{\Delta m^2_{\text{atm}} L}{4E}\right),$$

$$P_{\nu_\mu \to \nu_\tau} \approx \sin^2(2\theta_{23}) \cos^2(\theta_{13}) \ \sin^2\left(\frac{\Delta m^2_{\text{atm}} L}{4E}\right),$$

$$P_{\nu_e \to \nu_\mu} \approx \sin^2(2\theta_{13}) \sin^2(\theta_{23}) \ \sin^2\left(\frac{\Delta m^2_{\text{atm}} L}{4E}\right).$$

11.5 Consider neutrinos that are produced in the core of the sun, where $n_e \sim 3 \times 10^{25}$ cm^{-3}. Under these conditions, what is the probability that an electron neutrino will oscillate into a muon neutrino over a baseline of 3×10^8 cm? At what energy do matter effects cause these neutrinos to oscillate resonantly?

12

Baryogenesis and Leptogenesis

One of the biggest mysteries in modern cosmology is the question of how our universe came to contain so much more matter than antimatter. There have been many proposals for how this asymmetry may have been generated in the early universe, and this chapter describes some of the most prominent of these ideas. I begin by reviewing the matter-antimatter asymmetry itself and describing the criteria—known as the Sakharov conditions—that must be satisfied by any theory that can explain this feature of our universe. I then go on to discuss several classes of theories for how this asymmetry may have been generated, including baryogenesis in the context of grand unified theories, leptogenesis, and electroweak baryogenesis.

12.1 The matter-antimatter asymmetry

While our universe contains significant quantities of baryonic matter (nucleons, atoms, etc.), it harbors almost no antimatter. Furthermore, the trace abundances of antimatter that do exist are not primordial in origin, but were formed rather recently in the collisions of cosmic rays. By the time of Big Bang nucleosynthesis, any traces of primordial antimatter that had once existed in our universe had disappeared.

The laws of physics, as we currently understand them, offer no explanation for this observed asymmetry between the abundances of matter and antimatter. Every interaction that is experienced by matter is mirrored by its antimatter counterpart. Furthermore, baryon number and lepton number are each almost perfectly conserved in the Standard Model, at least at temperatures below the electroweak phase transition. Even at higher temperatures, the difference between the total baryon number and lepton number, $B - L$, is conserved. Nothing we know about our universe would have made it possible for a significant matter-antimatter asymmetry to have arisen from a matter-antimatter symmetric state.

In light of this puzzle, one might be tempted to push this problem back to the earliest of times, in an effort to treat the matter-antimatter asymmetry as an initial condition of our universe. If this were possible, it could resolve the problem at hand. After all, for the same reason that we would not expect a matter-antimatter asymmetry to arise from a symmetric state, any asymmetry between matter and antimatter should remain in place once present.

The problem with this approach is that our best theories do not give us any reason to suspect that our universe was initially in an asymmetric state. During inflation, for example, there was no baryon or lepton number present, and thus no asymmetry between matter and antimatter. Instead, this asymmetry seems to have somehow been generated after inflation but before the onset of Big Bang nucleosynthesis.

At this point, you might be asking yourself just how confident we are that our universe doesn't contain any significant quantities of antimatter. While it is clear that our local universe is overwhelmingly populated by matter, it is less obvious that distant regions of our universe are not instead inhabited by antimatter galaxies, consisting of antimatter stars, planets, gas, and dust. After all, in isolation, the light emitted from an antimatter star or galaxy would be indistinguishable from that emitted from an ordinary star or galaxy made of matter. The problem with this hypothesis is that stars and galaxies, whether made of matter or antimatter, do not exist in perfect isolation. In any realistic scenario featuring large quantities of antimatter, particles of antimatter would inevitably come into contact with matter, producing distinctive annihilation signals, such as 511 keV photons from electron-positron annihilation, $e^+e^- \to \gamma\gamma$. Such signatures of antimatter stars or galaxies have not been forthcoming, leading us to conclude that our observable universe is, in fact, overwhelmingly dominated by matter.

Another argument for the lack of antimatter in our universe comes from considerations regarding the origin of nucleons in the early universe. If equal numbers of nucleons and antinucleons (or quark and antiquarks, prior to the QCD phase transition) had been present, we can calculate how the abundances of these particles would have evolved in the early universe. Similar to our treatment of dark matter particles in Sec. 9.5, we can write

$$\frac{dn_{N+\bar{N}}}{dt} = -3Hn_{N+\bar{N}} - \frac{1}{2}\langle\sigma\, v\rangle\left[n^2_{N+\bar{N}} - (n^{\rm Eq}_{N+\bar{N}})^2\right], \tag{12.1}$$

where $\langle\sigma\, v\rangle \sim m_\pi^{-2} \sim 10^{-15}\,{\rm cm}^3/{\rm s}$ is the thermally averaged cross section for nucleon-antinucleon annihilation. The combined abundance of nucleons and antinucleons freezes out of equilibrium when the Hubble term in the above Boltzmann equation becomes larger than the annihilation term (see Eq. 9.17). This takes place at a temperature T_f, given by

$$\begin{aligned}\frac{m_N}{T_f} &= -\ln\left[\frac{8\pi^3 g_\star^{1/2}(T_f)\,T_f^{1/2}}{3\sqrt{10}\,g_N m_N^{3/2} M_{\rm Pl}\,\langle\sigma\, v\rangle}\right] \\ &\longrightarrow T_f \sim 20\,{\rm MeV},\end{aligned} \tag{12.2}$$

leading to the following abundance of nucleons and antinucleons (see Eq. 9.18):

$$\begin{aligned}\rho_{N+\bar{N}} &\approx 1.5\times 10^{-16}\,{\rm GeV/cm}^3 \times \left(\frac{10^{-15}\,{\rm cm}^3/{\rm s}}{\langle\sigma\, v\rangle}\right)\left(\frac{10}{g_\star(T_f)}\right)^{1/2}\left(\frac{20\,{\rm MeV}}{T_f}\right) \\ &\sim 6\times 10^{-10}\,\rho_b.\end{aligned} \tag{12.3}$$

In summary, if nucleons and antinucleons had existed in equal quantities in the early universe, they would have annihilated with each other so efficiently as to leave behind an abundance that is more than a billion times smaller than the measured baryon density. Thus the very fact that protons and neutrons exist in significant quantities in our universe requires that there must have been a primordial asymmetry between matter and antimatter.

The matter-antimatter asymmetry of our universe is sometimes quantified in terms of the net number density of baryons, $n_B = n_b - n_{\bar{b}}$, relative to the number density of photons,

$$\eta \equiv \frac{n_B}{n_\gamma} = \frac{n_b - n_{\bar{b}}}{n_\gamma}. \tag{12.4}$$

While the value of η has remained constant over most of cosmic history, entropy dumps (see Sec. 4.2) caused this quantity to evolve in the early universe. With this in mind, it can be useful to consider the comoving baryon number density, $n_B a^3$, which is conserved so long as no net quantities of baryon number are created or destroyed. From entropy conservation, we can relate $n_B \propto a^{-3} \propto g_{\star,S}\, T^3$, which in turn is proportional to the total entropy density, $s = 2\pi^2 g_{\star,S} T^3/45$. We can thus quantify the baryon asymmetry in terms of the ratio of the net baryon density to the entropy density:

$$\begin{aligned} Y_B \equiv \frac{n_B}{s} &= \frac{n_b - n_{\bar{b}}}{s} \\ &= \frac{45 n_{B,0}}{2\pi^2}, \end{aligned} \tag{12.5}$$

where $n_{B,0}$ is the present value of n_B. Note that for $T \ll$ MeV (corresponding to $g_{\star,S} \approx 3.909$), this quantity is related to the baryon-to-photon ratio by $Y_B \approx 0.14\,\eta$.

In the absence of any net baryon number changing processes, Y_B remains constant as the universe expands and cools. In principle, of course, one could imagine processes that create or destroy net baryon number, causing Y_B to evolve with time. That being said, measurements of the cosmic microwave background and the primordial light element abundances each yield $Y_B \approx 8.8 \times 10^{-11}$, indicating that the current magnitude of the baryon asymmetry was in place by the era of Big Bang nucleosynthesis. This forces us to conclude that our universe's baryon asymmetry was generated early in cosmic history, during the first second or so after the Big Bang.

12.2 The Sakharov conditions

Although we do not know precisely how the baryon asymmetry of our universe was generated, the process or processes that were responsible for this must have satisfied what are known as the Sakharov conditions [115]:

1. The violation of baryon number conservation.

2. The violation of both charge (C) and charge-parity (CP) symmetries.

3. Departure from thermal equilibrium.

From among these conditions, the first is perhaps most obvious. After all, any closed system that starts out with zero net baryon number will remain without any net baryon number so long as baryon number is conserved. Sakharov's second condition is more subtle, but comes down to the fact that if charge and charge-parity symmetries are each exactly enforced, then the rate for any process that generates baryon number will be equal to that of the corresponding antimatter generating process. For example, imagine that we wanted to produce net baryon number through the baryon number violating process, $e^+ + \nu_e \to p + \gamma$. If C and CP are both good symmetries, this process will be accompanied by an equal rate of $e^- + \bar{\nu}_e \to \bar{p} + \gamma$, leading to no net change in baryon number.

The third Sakharov condition reflects the fact that all processes are reversible under the conditions of thermal equilibrium, requiring any baryon number changing process to occur at a rate that is equal to its reverse. Considering the same interaction as in the previous example, the rate for the process $e^+ + \nu_e \to p + \gamma$ would be identical to that for $p + \gamma \to e^+ + \nu_e$ under the conditions of thermal equilibrium. Another way to see why this must be true is to recognize that, in chemical equilibrium, baryon number violation would require all particles with baryon number to have zero chemical potential. In equilibrium with no chemical potential, the number density of a baryon number carrying species will be equal to that of its antiparticle, corresponding to a state with no zero net baryon number.

The first two Sakharov conditions are satisfied within the context of the Standard Model, at least to an extent. Non-perturbative processes involving electroweak sphalerons can create or destroy baryon number, while preserving the difference of baryon and lepton number, $B - L$ [14, 15]. Furthermore, the symmetries of C and CP are each violated by the weak interactions. These features, however, are insufficient to generate the observed baryon asymmetry. Physics beyond the Standard Model is required to explain how our universe came to contain so many baryons and so few antibaryons.

12.3 Baryogenesis in grand unified theories

Grand unified theories (GUTs) are among the most well-motivated and widely studied extensions of the Standard Model. As we will see in this section, they also provide a framework in which a baryon asymmetry could plausibly have been generated [116, 117, 118, 119, 120]. In particular, GUTs generically predict the existence of heavy particles whose decays violate the conservation of baryon number, as well as the symmetries of C and CP. If these particles decayed after falling out of equilibrium in the early universe, they could have produced the observed baryon asymmetry of our universe.

Although quite empirically successful, the Standard Model of particle physics is far from simple. In particular, the Standard Model is based on the product of three gauge groups, $SU(3)_C \times SU(2)_L \times U(1)_Y$, each with its own gauge coupling. Furthermore, the Standard Model has a total of 19 free parameters, whose values are simply determined from data. It has long been imagined that there might exist a simpler and more elegant theory from which these parameters could be derived. GUTs attempt to do this by embedding the structure and symmetries of the Standard Model within a larger gauge group.

The gauge group of the Standard Model has rank 4, corresponding to two for $SU(3)_C$ and one for each of $SU(2)_L$ and $U(1)_Y$. As a consequence, any GUT must be based on a gauge group of at least this rank. With this in mind, the rank 4 group, $SU(5)$, is the simplest gauge group that contains all of the symmetries of the Standard Model, thus representing the minimal choice for grand unification [121]. In particular, the **10** and $\bar{\mathbf{5}}$ representations of $SU(5)$ (the smallest two irreducible representations of this group) have the following structure:

$$\mathbf{10}: \begin{pmatrix} 0 & u_b^c & -u_g^c & u_r & d_r \\ -u_b^c & 0 & u_r^c & u_g & d_g \\ u_g^c & -u_r^c & 0 & u_b & d_b \\ -u_r & -u_g & -u_b & 0 & e^c \\ -d_r & -d_g & -d_b & -e^c & 0 \end{pmatrix}, \qquad \bar{\mathbf{5}}: \begin{pmatrix} d_r^c \\ d_g^c \\ d_b^c \\ e \\ -\nu \end{pmatrix}, \tag{12.6}$$

where the subscripts label QCD color (b, g, r), and the superscript, c, denotes the charge conjugate of the state in question. From this, we see that the **10** and $\bar{\mathbf{5}}$ representations of $SU(5)$ contain precisely the fermionic particle content of one generation of the Standard Model—no more and no less, and with all of the correct charge assignments. From this perspective, it is easy to see why so many physicists have found GUTs based on $SU(5)$ to be so compelling. Furthermore, since each of these representations contain both quarks and leptons, gauge bosons will exist in these theories that can transform quarks into leptons and vice versa, and thus can destroy and create baryon number. For this reason, Sakharov's first condition is easily satisfied within the context of GUTs.

As an alternative to $SU(5)$, many GUTs have been based on the rank 5 gauge group, $SO(10)$. One **16** representation of this group contains all of the fermions in one generation of the Standard Model, along with one Standard Model singlet that can be identified with a right-handed neutrino, allowing for the generation of light neutrino masses through the seesaw mechanism (see Sec. 11.1) [122, 123]. Other GUTs have been based on a variety of larger symmetry groups, including E_6.

The number of gauge bosons associated with a given gauge group is determined by the number of its generators. The $SU(N)$ group, for example, has $N^2 - 1$ generators. Thus whereas the gauge group of QCD, $SU(3)$, features $3^2 - 1 = 8$ gluons, GUTs based on $SU(5)$ predict the existence of 24 different gauge bosons. These consist of eight gluons, four electroweak gauge bosons (corresponding to the photon, Z, W^+, and W^-), and 12 other gauge bosons represented by the symbols X and Y.[1] These extra gauge bosons form an $SU(3)_C$ triplet and an $SU(2)_L$ doublet, and, unlike the gauge bosons of the Standard Model, their interactions can transform quarks into leptons and vice versa, thereby violating the conservation of both baryon and lepton number.

These baryon number violating X and Y bosons can mediate processes that allow the proton to decay, such as $p \to \pi^0 e^+$. Although the precise rate for this process is somewhat model-dependent, it is roughly given by

$$\begin{aligned}\Gamma_{p\to\pi^0 e^+} &\approx \frac{g_{\rm GUT}^4}{16\pi^2}\frac{m_p^5}{M_{X,Y}^4} \\ &\approx (1.7\times 10^{34}\,{\rm yr})^{-1}\times\left(\frac{g_{\rm GUT}}{0.04}\right)^4\left(\frac{3\times 10^{15}\,{\rm GeV}}{M_{X,Y}}\right)^4,\end{aligned} \tag{12.7}$$

where $g_{\rm GUT}$ is the strength of the GUT gauge coupling, and $M_{X,Y}$ are the masses of the X and Y bosons. Experiments including Super-Kamiokande have placed stringent constraints on the proton's lifetime, restricting $\tau_p > 1.67\times 10^{34}$ yr for the process $p \to \pi^0 e^+$ [124]. In light of this, we conclude that the baryon number violating gauge bosons must be very heavy, $M_{X,Y} \gtrsim 3\times 10^{15}\,{\rm GeV}\times(g_{\rm GUT}/0.04)$. In the simplest GUTs based on $SU(5)$, the breaking of this symmetry occurs at a temperature on the order of $\sim 5\times 10^{14}$ GeV. As the masses of the heavy gauge bosons are expected to be near this symmetry breaking scale, such theories appear to be ruled out by the constraints on proton decay. If supersymmetric particles (see Sec. 17.1) exist with masses at or around the TeV scale, however, then the Standard Model gauge couplings will instead evolve to a common value of $g_1 \approx g_2 \approx g_3 \approx g_{\rm GUT} \approx 0.04$ at a somewhat higher energy scale of $\sim 10^{16}$ GeV [125, 126]. In such a scenario, we expect $M_{X,Y} \sim 10^{16}$ GeV, which is consistent with all existing constraints from proton decay.

[1] The symbols X and Y are used to denote gauge bosons with electric charges of $Q_X = -1/3$ and $Q_Y = -4/3$, respectively.

In order to break both electroweak symmetry and the GUT gauge symmetry, as well as to give masses to the quarks and leptons, GUTs must also include Higgs bosons. The Higgs bosons that break the GUT symmetry are contained in the adjoint representation of the GUT group, while those that facilitate electroweak symmetry breaking depend on the details of the model. For example, in $SU(5)$, a **5** (or $\bar{\mathbf{5}}$) representation contains both an $SU(2)$ doublet, which includes the Higgs boson that breaks electroweak symmetry, as well as a color triplet [127, 128, 129, 130, 131]. Like the X and Y bosons, these color triplet Higgs bosons must be heavy ($\gtrsim 10^{11}$ GeV) in order to avoid inducing an unacceptable rate of proton decay [132].

In order to generate the observed baryon asymmetry in the context of a GUT, a large abundance of heavy X or Y bosons must be created in the early universe, which then fall out of equilibrium before decaying. This sequence of events can satisfy all three of Sakharov's conditions and lead to the observed magnitude of Y_B. In particular, as the decays of X and Y each violate the conservation of baryon number, the first condition is automatically satisfied. Since the weak interactions are known to violate the symmetries of both C and CP, it is very plausible that this will also be exhibited by the heavy GUT bosons. Finally, by decaying after falling out of equilibrium, the processes that generate baryon number will be irreversible, thereby satisfying Sakharov's third condition.

The heavy boson population will remain in chemical equilibrium so long as the processes that change the net number of these particles occur at a rate that is faster than the rate of Hubble expansion. This includes the rates for decay, inverse decay, and production or destruction via scattering. In practice, it is generally the decay rate of the heavy particles that is the most rapid, and thus the condition for the departure from equilibrium can be written as $\Gamma_X \lesssim H$. Estimating $\Gamma_X \sim g_{\rm GUT}^2 M_X/4\pi$, the ratio of these rates is given by

$$\begin{aligned}\frac{\Gamma_X}{H} &\sim \frac{g_{\rm GUT}^2 M_X}{4\pi}\left(\frac{3M_{\rm Pl}^2}{8\pi\rho}\right)^{1/2} \\ &\sim \left(\frac{10^{15}\,{\rm GeV}}{T}\right)^2\left(\frac{106.75}{g_\star(T)}\right)^{1/2}\left(\frac{g_{\rm GUT}}{0.04}\right)^2\left(\frac{M_X}{10^{16}\,{\rm GeV}}\right),\end{aligned} \tag{12.8}$$

where in the last line I have used $\rho \sim \pi^2 g_\star(T)\, T^4/30$, thereby assuming that the energy density was dominated by Standard Model particle species in thermal equilibrium. From this expression, we see that at very early times, $T \gtrsim 10^{15}$ GeV, the X population was decaying at a rate that was slower than the rate of Hubble expansion, providing a window during which such decays could plausibly have satisfied Sakharov's third condition.

In some GUTs, the baryon asymmetry could instead be generated through the out-of-equilibrium decays of the scalar Higgs bosons that are part of the color triplet. In this case, we replace $g_{\rm GUT}$ in Eq. 12.8 with the appropriate Yukawa coupling, y. If this Yukawa coupling is small, $y \ll g_{\rm GUT}$, the scalar Higgs bosons could easily decay after falling out of equilibrium, even if they are as light as the minimum value allowed by proton decay constraints, $M \sim 10^{11}$ GeV. Although we will focus here on the decays of GUT gauge bosons, much of what we will discuss in this section could also be applied to the case of baryon number violating scalar decays.

During any period in which the Hubble rate exceeds the X decay rate, the number density of these particles will evolve approximately as $n_X \propto a^{-3}$, and will not feature the exponential suppression that characterizes the equilibrium number density of a nonrelativistic species. This means that the number density of these particles will be comparable to that of photons prior to their decays, $n_X \sim n_\gamma$. Note that while processes such as inverse decays could

potentially impact the value of n_X, the rate of such interactions is generally much smaller than the Hubble rate when $T \lesssim M_X$.

We are now in a position to estimate the magnitude of the baryon asymmetry that will be generated through out-of-equilibrium X decays:

$$\begin{aligned} Y_B \equiv \frac{n_B}{s} &\sim \frac{45\epsilon_X n_X}{2\pi^2 g_{\star,S}(T) T^3} \\ &= \frac{45\zeta(3)\epsilon_X n_X}{\pi^4 g_{\star,S}(T)\, n_\gamma} \\ &\approx 8.8 \times 10^{-11} \times \left(\frac{\epsilon_X}{1.7 \times 10^{-8}}\right)\left(\frac{n_X}{n_\gamma}\right)\left(\frac{106.75}{g_{\star,S}(T)}\right), \end{aligned} \tag{12.9}$$

where ϵ_X is the average amount of net baryon number that is generated per decay. The quantity n_X in this expression is evaluated prior to the decays, while n_γ and $g_{\star,S}$ are each evaluated immediately after. Comparing this to the measured value of Y_B, we find that these decays could produce the observed baryon asymmetry even if only a small amount of baryon number is generated in an average decay, $\epsilon_X \sim 2 \times 10^{-8}$.

The value of ϵ_X can be expressed in terms of the branching fractions of X and the total baryon number of the corresponding final states:

$$\epsilon_X = \sum_f B_f \, \frac{\Gamma(X \to f) - \Gamma(\bar{X} \to \bar{f})}{\Gamma_X}. \tag{12.10}$$

The sum in this expression is performed over the relevant final states, each with a baryon number, B_f (and where the state $\bar{f}$ carries baryon number, $-B_f$). $\Gamma(X \to f)$ is the partial width to final state, f, and Γ_X is the total decay width. As it turns out, X and Y bosons can each decay into a pair of quarks ($B_f = 2/3$), or into an antiquark and an antilepton ($B_f = -1/3$). It is precisely the fact that these bosons have multiple decay channels, with different amounts of total baryon number, that enables baryon number to be generated in this class of theories.

In contrast to baryon number, baryon-minus-lepton number is conserved in some GUTs, in particular in those based on $SU(5)$ (notice that a pair of quarks has the same total baryon-minus-lepton number as an antiquark and an antilepton, $B - L = 2/3$). In GUTs based on gauge groups larger than $SU(5)$, however, baryon-minus-lepton is not necessarily conserved. In $SO(10)$ GUTs, for example, $B - L$ is spontaneously broken, making it possible for the decays of heavy bosons to generate a net asymmetry not only in baryon number, but in $B - L$ as well. This will be an important feature when we later consider the influence of electroweak sphalerons on a baryon asymmetry.

Eq. 12.10 makes Sakharov's requirement of CP violation more explicit. In particular, if CP is a good symmetry of the theory, then $\Gamma(X \to f) = \Gamma(\bar{X} \to \bar{f})$, leading to $\epsilon_X = 0$. To see what features a theory must possess in order to predict a nonzero value of ϵ_X, consider an interaction of the form $\mathcal{L} = gXi_2^\dagger i_1 + \text{h.c.}$, where i_1 and i_2 represent fermions. Through this interaction vertex, the tree-level decays of the X proceed with widths given by $\Gamma(X \to \bar{i}_1 i_2) = |g|^2 I_X$ and $\Gamma(\bar{X} \to i_1 \bar{i}_2) = |g|^2 I_X$, where I_X is a factor obtained from the relevant phase space integrals. Since this automatically leads to $\Gamma(X \to \bar{i}_1 i_2) = \Gamma(\bar{X} \to i_1 \bar{i}_2)$, we conclude that it is not possible to generate a baryon asymmetry in a theory with only one heavy boson.

Fortunately, there are multiple heavy bosons in most GUTs, making it possible to obtain a finite value for ϵ_X. Consider a theory with two such bosons, X and Y, with interactions

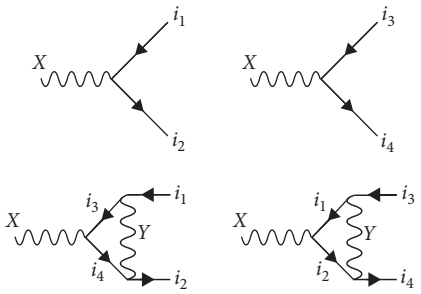


Figure 12.1 Tree-level and one-loop diagrams for the decay of a heavy GUT boson.

that are described by the following Lagrangian:

$$\mathcal{L} = g_1 X i_2^\dagger i_1 + g_2 X i_4^\dagger i_3 + g_3 Y i_1^\dagger i_3 + g_4 Y i_2^\dagger i_4 + \text{h.c.} \tag{12.11}$$

The partial widths of the X to the $\bar{i}_1 i_2$ and $i_1 \bar{i}_2$ final states in this case are given by:

$$\begin{aligned}\Gamma(X \to \bar{i}_1 i_2) &\approx |g_1|^2 I_X + g_1\, g_2^\star\, g_3\, g_4^\star\, I_{XY} + (g_1\, g_2^\star\, g_3\, g_4^\star\, I_{XY})^\star, \\ \Gamma(\bar{X} \to i_1 \bar{i}_2) &\approx |g_1|^2 I_X + g_1^\star\, g_2\, g_3^\star\, g_4\, I_{XY} + (g_1^\star\, g_2\, g_3^\star\, g_4\, I_{XY})^\star.\end{aligned} \tag{12.12}$$

The first term in each of these expressions arises entirely from a tree-level diagram, and together they contribute zero to ϵ_X. The second and third terms, in contrast, result from the interference between the tree-level and one-loop processes (see Figure 12.1). Once again, I_{XY} is a phase-space factor that, in general, can be complex. Taking the difference between these terms, we find:

$$\begin{aligned}\Gamma(X \to \bar{i}_1 i_2) - \Gamma(\bar{X} \to i_1 \bar{i}_2) &\approx (g_1\, g_2^\star\, g_3\, g_4^\star - g_1^\star\, g_2\, g_3^\star\, g_4)\, I_{XY} \\ &\quad + (g_1^\star\, g_2\, g_3^\star\, g_4 - g_1\, g_2^\star\, g_3\, g_4^\star)\, I_{XY}^\star \\ &= 2i\,\text{Im}(g_1\, g_2^\star\, g_3\, g_4^\star)\, I_{XY} + 2i\,\text{Im}(g_1^\star\, g_2\, g_3^\star\, g_4)\, I_{XY}^\star \\ &= -4\,\text{Im}(I_{XY})\,\text{Im}(g_1\, g_2^\star\, g_3\, g_4^\star).\end{aligned} \tag{12.13}$$

In a similar fashion, we can calculate the result for the other decay mode of the X boson:

$$\Gamma(X \to \bar{i}_3 i_4) - \Gamma(\bar{X} \to i_3 \bar{i}_4) \approx 4\,\text{Im}(I_{XY})\,\text{Im}(g_1\, g_2^\star\, g_3\, g_4^\star). \tag{12.14}$$

Combining these results with Eq. 12.10, we obtain the following:

$$\begin{aligned}\epsilon_X &= B_{\bar{i}_1 i_2}\left[\frac{\Gamma(X \to \bar{i}_1 i_2) - \Gamma(\bar{X} \to i_1 \bar{i}_2)}{\Gamma_X}\right] + B_{\bar{i}_3 i_4}\left[\frac{\Gamma(X \to \bar{i}_3 i_4) - \Gamma(\bar{X} \to i_3 \bar{i}_4)}{\Gamma_X}\right] \\ &\approx \frac{4\,\text{Im}(I_{XY})\,\text{Im}(g_1\, g_2^\star\, g_3\, g_4^\star)}{\Gamma_X}\left[-B_{\bar{i}_1 i_2} + B_{\bar{i}_3 i_4}\right].\end{aligned} \tag{12.15}$$

An analagous calculation for the decays of the Y boson yields $\epsilon_Y = -\epsilon_X$.

From this exercise, we have learned that to generate a net baryon asymmetry, we need not only one, but two heavy bosons, X and Y. Furthermore, to accommodate the requirement of CP violation, both I_{XY} and $g_1 g_2^\star g_3 g_4^\star$ must be complex. If $\mathrm{Im}(I_{XY}) \sim I_X, I_Y$, and the final state fermions carry order 1 baryon number, this leads to $|\epsilon_{X,Y}| \sim 4\mathrm{Im}(g_1\, g_2^\star\, g_3\, g_4^\star)/(|g_1|^2 + |g_2|^2)$. For couplings of magnitude $g_{\rm GUT} \sim 0.04$, this quantity is expected to be no larger than $\epsilon_{X,Y} \sim 0.003$, and potentially much smaller if $g_1 g_2^\star g_3 g_4^\star$ is not mostly imaginary. Fortunately, only a small value of $\epsilon_{X,Y} \sim 2 \times 10^{-8}$ is needed to accomplish successful baryogenesis.

More generally speaking, GUTs are but one of the possible frameworks in which baryogenesis could occur through out-of-equilibrium and baryon number violating decays. In fact, most models of GUT baryogenesis face significant challenges, and this has led many cosmologists to direct their attention toward other ideas for how the baryon asymmetry may have been generated in the early universe. In particular, models of GUT baryogenesis generally require a large abundance of very heavy gauge bosons ($M_{X,Y} \sim 10^{16}$ GeV) or heavy scalars ($M \gtrsim 10^{11}$ GeV) to have been produced in the early universe, and it is not at all clear that our universe was ever hot enough to produce such particles thermally (see Chapter 13). While mechanisms have been proposed for the efficient production of massive particles prior to post-inflationary reheating [133, 134, 135, 136, 137], this issue has long been something of a thorn in the side of traditional theories of GUT baryogenesis.

Another significant challenge faced by theories of GUT baryogenesis has to do with the influence of electroweak sphalerons on any baryon asymmetry that may have been generated at early times. In particular, at temperatures below $T \sim 10^{12}$ GeV but above the electroweak phase transition, sphaleron transitions are predicted to efficiently change the amount of baryon-plus-lepton number ($B + L$) that is present in the thermal bath, driving this quantity toward zero and thus erasing any asymmetry that may have been present at high temperatures [138]. This effect of sphaleron washout will be the topic of the following section.

12.4 Sphalerons

In the Standard Model, the structure of the vacuum is nontrivial. In particular, there exist many different vacua with different values of both baryon number and lepton number. Any transitions that take place between these different vacuum states will thus result in the creation or destruction of baryon and lepton number. To be more specific, such transitions change the total baryon number and the total lepton number by multiples of 3, while keeping the quantity $B - L$ constant. A vacuum transition might, for example, convert nine antiquarks into three leptons, thereby increasing the baryon number and lepton number each by 3, while keeping the value of $B - L$ unchanged (see Figure 12.2).

Due to the large potential barriers that separate these vacua, transitions of this kind are extremely rare in the current era. While tunneling between vacuum states is not strictly prohibited, the corresponding rates are suppressed by a factor of $e^{16\pi^2/g_2^2} \sim 10^{161}$ [14, 15]. This fact makes it entirely unsurprising that transitions between Standard Model vacua have never been directly observed. At temperatures above the electroweak phase transition, however, such transitions can instead proceed through thermal fluctuations over the barrier, without the large degree of suppression that comes with quantum tunneling. These transitions are fundamentally non-perturbative in nature, and thus cannot be represented by Feynman diagrams. Transitions between vacua which occur through tunneling are called instantons, while the unsuppressed high-temperature transitions are known as electroweak sphalerons.

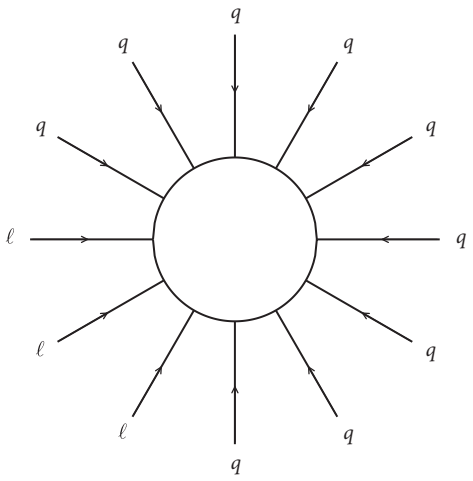


Figure 12.2 The effective baryon number and lepton number violating vertex associated with electroweak sphalerons. Such interactions change the total amount of baryon number and lepton number, $\Delta B = \Delta L = \pm 3$, while conserving the difference of these quantities, $\Delta(B-L) = 0$.

Sphaleron transitions can be thought of as occurring through the minimum energy path that connects two vacua with a difference of 3 in baryon number, $\Delta B = 3$. Counterintuitively (at least to me), a sphaleron is a spatially extended entity, with a scale of roughly $l_{\rm sp} \sim 1/g_2 v \sim 10^{-16}$ cm. For this reason, a sphaleron can act coherently across a large number of particles in a high density environment. In the early universe, the number of particles within a sphere of radius $l_{\rm sp}$ was on the order of $N \sim n\, l_{\rm sp}^3 \sim T^3/(g_2 v)^3 \sim 300 \times (T/{\rm TeV})^3$. This large number of particles can work against the suppression that would normally be associated with transitions between vacuum states. A rough estimate for the amplitude of such a transition can be written as

$$|\mathcal{M}|^2 \sim \left(\frac{4\pi}{g_2^2}\right)^N e^{-16\pi^2/g_2^2}, \tag{12.16}$$

where the second factor is the exponential suppression associated with low-temperature vacuum transitions. This suppression can be overcome, allowing for efficient transitions between vacua, if $N \gtrsim 16\pi^2/g_2^2 \ln(4\pi/g_2^2) \sim 100$, corresponding to $T \gtrsim \mathcal{O}({\rm TeV})$.

At temperatures above the electroweak phase transition, it is common to parameterize the rate of sphaleron transitions per volume as follows [139, 140]:

$$\Gamma_{\rm sp} \sim 25 \left(\frac{g_2^2}{4\pi}\right)^5 T^4. \tag{12.17}$$

In principle, such transitions can occur in both directions, either increasing or decreasing the value of $B+L$. Each choice of vacuum, however, is associated with a different amount of free energy, and thus transitions toward larger or smaller values of $B+L$ will not generally occur at the same rates. To estimate the magnitude of this effect, we can write the density of $B+L$ in terms of a chemical potential, $n_{B+L} \sim \mu_{B+L}\, T^2$, and thus estimate the density of free energy as $\rho_F \sim \mu_{B+L}^2 T^2 \sim n_{B+L}^2/T^2$ for $\mu_{B+L} \gtrsim T$. The main point here is that since the free energy is minimized for the smallest values of n_{B+L}^2, the net value of $B+L$ tends to be driven toward zero whenever sphaleron transitions are efficient.

The evolution of the total density of baryon number can be described by the following Boltzmann equation [141]:

$$\begin{aligned}\frac{dn_B}{dt} &= -3Hn_B + 3\left(\Gamma_{\rm sp,+} - \Gamma_{\rm sp,-}\right) \\ &= -3Hn_B - \frac{\Gamma_{\rm sp}\,\mathcal{M}\,n_B}{T^3},\end{aligned} \tag{12.18}$$

where $\Gamma_{\rm sp,+}$ and $\Gamma_{\rm sp,-}$ represent the rates for transitions to a state with higher or lower $B+L$, respectively. The quantity $\mathcal{M}$ is related to the number of fermionic degrees of freedom that are present in thermal equilibrium. In the Standard Model, $\mathcal{M} = 39/2$ at high temperatures. Using Eq. 12.17 for the sphaleron rate, the ratio of the sphaleron and Hubble terms is given by the following:

$$\frac{\Gamma_{\rm sp}\,\mathcal{M}\,n_B/T^3}{3Hn_B} \sim \left(\frac{10^{12}\,{\rm GeV}}{T}\right)\left(\frac{106.25}{g_\star}\right)^{0.5}. \tag{12.19}$$

From this, we see that at temperatures between $T \sim 10^{12}$ GeV and the electroweak phase transition, the effects of sphaleron washout are very efficient, driving the abundances of baryon number and lepton number each toward zero, while maintaining a fixed value of $B-L$.

Electroweak sphalerons significantly impact how we think about baryogenesis. In particular, if a baryon asymmetry was generated through out-of-equilibrium decays at a temperature above the electroweak phase transition, electroweak sphalerons would rapidly reduce the density of baryon-plus-lepton number, $B+L$, potentially erasing any asymmetry that may have been present at high temperatures [138]. This is particularly problematic for GUTs based on $SU(5)$, which feature heavy bosons whose decays produce a net abundance of $B+L$, but no net $B-L$. In this case, any baryon asymmetry that is produced will be efficiently washed out well before the electroweak phase transition. One way to avoid this problem is to consider GUTs, such as those based on $SO(10)$, which contain heavy bosons that produce a nonzero value of $B-L$ in their decays, which is not eroded by sphalerons.

Before moving on, let's consider how much net baryon number and net lepton number will remain after interactions in the early universe (including those involving electroweak sphalerons) have restored the plasma to a state of thermal equilibrium (i.e., after sphaleron washout). To this end, we can use the relevant interactions to relate the chemical potentials of the various Standard Model particle species (see Sec. 4.3). Assuming that the chemical potentials of the fermions are the same for all three generations ($\mu_{e_L} = \mu_{\mu_L} = \mu_{\tau_L}$, for example), we can place the following constraints on these quantities [142]:

1. Through electroweak sphalerons, fermions can interact in such a way that leads to the following relationship between the $SU(2)$ doublet quarks and leptons:

$$N_c \times (\mu_{u_L} + \mu_{d_L}) + \mu_{e_L} + \mu_{\nu_L} = 0, \tag{12.20}$$

   where $N_c = 3$ is the number of colors.

2. QCD instanton processes allow for interactions between the $SU(2)$ doublet and $SU(2)$ singlet quarks, leading to:

$$\mu_{u_L} + \mu_{d_L} - \mu_{u_R} - \mu_{d_R} = 0. \tag{12.21}$$

3. Hypercharge is conserved in the Standard Model, allowing us to write:

$$N_g \times \Big[N_c \times (Y_{u_L}\mu_{u_L} + Y_{u_R}\mu_{u_R} + Y_{d_L}\mu_{d_L} + Y_{d_R}\mu_{d_R}) + Y_{e_L}\mu_{e_L} + Y_{e_R}\mu_{e_R} + Y_{\nu_L}\mu_{\nu_L} \Big] + 4Y_H\mu_H = 0 \tag{12.22}$$

$$\longrightarrow 3\mu_{u_L} + 12\mu_{u_R} + 3\mu_{d_L} - 6\mu_{d_R} - 3\mu_{e_L} - 6\mu_{e_R} - 3\mu_{\nu_L} + 4\mu_H = 0,$$

where $N_g = 3$ is the number of generations.

4. Lastly, interactions of the Higgs boson with fermions give rise to:

$$\begin{aligned} \mu_{u_L} - \mu_H - \mu_{d_R} &= 0, \\ \mu_{d_L} - \mu_H - \mu_{d_R} &= 0, \\ \mu_{u_L} + \mu_H - \mu_{u_R} &= 0, \\ \mu_{d_L} + \mu_H - \mu_{u_R} &= 0, \\ \mu_{e_L} - \mu_H - \mu_{e_R} &= 0, \\ \mu_{\nu_L} - \mu_H - \mu_{e_R} &= 0. \end{aligned} \tag{12.23}$$

Solving this system of equations, and assuming that all three generations of fermions were in equilibrium, we can write all of the chemical potentials in terms of one free parameter. In terms of μ_{e_L}, for example, these quantities can be related as follows:

$$\begin{aligned} \mu_{e_R} &= \frac{3}{7}\mu_{e_L}, \\ \mu_{d_R} &= -\frac{19}{21}\mu_{e_L}, \\ \mu_{u_R} &= \frac{5}{21}\mu_{e_L}, \\ \mu_{u_L} = \mu_{d_L} &= -\frac{1}{3}\mu_{e_L}, \\ \mu_{\nu_L} &= \mu_{e_L}, \\ \mu_H &= \frac{12}{21}\mu_{e_L}. \end{aligned} \tag{12.24}$$

Taking the difference between the number densities of quarks and antiquarks given in Eq. 4.5, we can write the total density of baryon number (in the $T \gg m$ limit) as follows (see also Eq. 4.17):

$$\begin{aligned} n_B &= N_g \times \frac{1}{3} \times (n_{u_L} + n_{u_R} + n_{d_L} + n_{d_R} - n_{\bar{u}_L} - n_{\bar{u}_R} - n_{\bar{d}_L} - n_{\bar{d}_R}) \\ &= \frac{\mu_{u_L}T^2}{2} + \frac{\mu_{u_R}T^2}{2} + \frac{\mu_{d_L}T^2}{2} + \frac{\mu_{d_R}T^2}{2} \end{aligned} \tag{12.25}$$

$$= \frac{T^2}{2}\left(-\frac{1}{3}\mu_{e_L} + \frac{5}{21}\mu_{e_L} - \frac{1}{3}\mu_{e_L} - \frac{19}{21}\mu_{e_L}\right)$$

$$= -\frac{2}{3}\, T^2\, \mu_{e_L},$$

where the factor of 1/3 in the first line is the baryon number of an individual quark. Similarly, we can write the total density of lepton number in terms of the related chemical potentials:

$$\begin{aligned} n_L &= N_g \times (n_{e_L^-} + n_{e_R^-} + n_{\nu_L} - n_{e_L^+} - n_{e_R^+} - n_{\bar{\nu}_R}) \\ &= 3 \times \left(\frac{\mu_{e_L} T^2}{6} + \frac{\mu_{e_R} T^2}{6} + \frac{\mu_{\nu_L} T^2}{6} \right) \\ &= \frac{T^2}{2}\left(\mu_{e_L} + \frac{3}{7}\mu_{e_L} + \mu_{e_L} \right) \\ &= \frac{17}{14}\, T^2\, \mu_{e_L}. \end{aligned} \tag{12.26}$$

From these two expressions, we see that the magnitudes of the net baryon number and lepton number densities each evolve in proportion to $T^2\, \mu_{e_L}$, and thus in proportion to one another, with $n_B = (-28/51)\, n_L$. We can also present these relationships in terms of the $B - L$ density:

$$n_B = \frac{28}{79}\, n_{B-L} \quad \text{and} \quad n_L = -\frac{51}{79}\, n_{B-L}. \tag{12.27}$$

To summarize the results of this calculation, we have shown that once equilibrium is reached in the early universe, the densities of baryon number and lepton number will each be proportional to the density of $B - L$. Thus if baryon number and lepton number are generated in equal quantities in the early universe, as happens in GUTs based on $SU(5)$, electroweak sphalerons will drive both n_B and n_L to zero, washing out any asymmetries that may have once been present. In order for a baryon asymmetry to persist, we must either have a means of generating net $B - L$ (and not merely $B + L$), or produce a net abundance of baryon number at a point in time after the electroweak phase transition.

Although electroweak sphalerons can pose a challenge for some theories of baryogenesis, they can in other contexts provide an opportunity for explaining the baryon asymmetry of our universe. In particular, one can consider the class of scenarios known as leptogenesis, in which out-of-equilibrium decays create a net abundance of lepton number in the early universe, without any net baryon number. Sphaleron transitions can then efficiently convert this lepton number asymmetry into a mixture of baryon and lepton number, thereby generating a baryon asymmetry. Leptogenesis will the subject of the following section.

12.5 Baryogenesis via leptogenesis

Leptogenesis is perhaps the single most popular framework for how our universe's baryon asymmetry may have been generated. In many respects, the mechanism behind leptogenesis is similar to that of GUT baryogenesis. In particular, it relies on the out-of-equilibrium

and C and CP violating decays of very heavy particles. Unlike in GUT baryogenesis, however, these decays do not directly produce any net baryon number. Instead, they create lepton number, which is later converted into a baryon asymmetry through the effects of electroweak sphalerons [143].

Part of the appeal of leptogenesis is that many of its required elements are the same as those needed to explain the existence of Standard Model neutrino masses. Recall from Chapter 11 that the generation of neutrino masses requires interactions involving right-handed neutrinos. With right-handed neutrinos present in our theory, we can write down interaction terms which generate both Dirac and Majorana masses for the Standard Model neutrinos (for context, see Eq. 11.6):

$$\begin{aligned}\mathcal{L} &= -\frac{1}{2} m_D (\bar{\nu}_L N_R + \bar{N}_L^C \nu_R^C) - \frac{1}{2} m_L \, \bar{\nu}_L \nu_R^C - \frac{1}{2} m_R \, \bar{N}_L^C N_R + \text{h.c.} \\ &= -\frac{1}{2} (\bar{\nu}_L, \bar{N}_L^C) \begin{pmatrix} m_L & m_D \\ m_D & m_R \end{pmatrix} \begin{pmatrix} \nu_R^C \\ N_R \end{pmatrix} + \text{h.c.}\end{aligned} \tag{12.28}$$

Given the very small values of the Standard Model neutrino masses, the seesaw mechanism is particularly well motivated. More specifically, for the choice of $m_R \gg m_D$ and $m_L = 0$, one automatically finds themselves in a situation with a very light and mostly left-handed neutrino, and a very heavy and mostly right-handed neutrino, with masses given by $m_\nu \approx m_D^2/m_R$ and $m_N \approx m_R$, respectively. The Dirac mass is related to the vacuum expectation value of the Higgs field and an associated Yukawa coupling, $m_D = y_\nu v/\sqrt{2}$. In terms of this Yukawa coupling and the value of the light neutrino mass, the mass of the corresponding right-handed neutrino is given by $m_N \sim 3 \times 10^{13}\,\text{GeV} \times (y_\nu/0.1)\,(0.01\,\text{eV}/m_\nu)$.

Majorana masses play a key role in leptogenesis. Unlike the interactions that generate a Dirac mass, the Majorana mass terms correspond to interactions that change the total lepton number by two units. It is precisely through this interaction that the decays of the heavy right-handed neutrinos are able to generate a primordial lepton asymmetry. In particular, the heavy right-handed neutrinos can decay into either a neutrino and a Higgs boson, or into an antineutrino and a Higgs boson. In either case, these decays create net lepton number, and thus clearly violate the conservation of this quantity.

Much as with GUT baryogenesis, it is straightforward to show that the right-handed neutrinos will decay after falling out of equilibrium. The rate for these decays is approximately given by:

$$\begin{aligned}\Gamma_N &\approx \frac{y_\nu^2 \, m_N}{16\pi} \\ &\approx \frac{m_\nu m_N^2}{8\pi \, v^2}.\end{aligned} \tag{12.29}$$

Comparing this to the rate of Hubble expansion, we find

$$\begin{aligned}\frac{\Gamma_N}{H} &\sim \frac{m_\nu m_N^2}{8\pi \, v^2} \left(\frac{3M_{\text{Pl}}^2}{8\pi\rho}\right)^{1/2} \\ &\sim \left(\frac{2 \times 10^{10}\,\text{GeV}}{T}\right)^2 \left(\frac{106.75}{g_\star(T)}\right)^{1/2} \left(\frac{m_\nu}{0.01\,\text{eV}}\right) \left(\frac{m_N}{10^{10}\,\text{GeV}}\right)^2,\end{aligned} \tag{12.30}$$

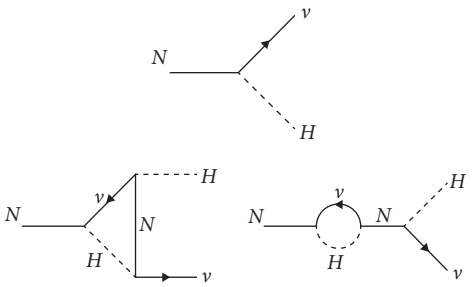


Figure 12.3 Tree-level and one-loop diagrams for the decay of a heavy right-handed neutrino.

thus demonstrating that at early times, when $T \gtrsim 10^{10}\,\text{GeV} \times (m_N/10^{10}\,\text{GeV})$, the right-handed neutrinos were decaying out of equilibrium, providing a window during which Sakharov's third condition could have been satisfied.

Just as we found in GUT baryogenesis, the tree-level decays of the right-handed neutrinos do not allow for any CP violation, since $\Gamma(N \to \nu H^c) = \Gamma(\bar{N} \to \bar{\nu} H)$, leading to $\epsilon_N = 0$. In order for CP to be violated in these decays, we must again rely on the interference terms between the tree-level and loop-level diagrams (see Figure 12.3). These interference terms lead to

$$\epsilon_{N_j} = \frac{\Gamma(N_j \to \nu H) - \Gamma(N_j \to \bar{\nu}\bar{H})}{\Gamma(N_j \to \nu H) + \Gamma(N_j \to \bar{\nu}\bar{H})} \tag{12.31}$$

$$\approx -\frac{1}{8\pi} \sum_{i \neq j} \frac{\text{Im}[(y_\nu^\dagger y_\nu)_{ij}^2]}{(y_\nu^\dagger y_\nu)_{jj}} f\left(\frac{m_{N_i}^2}{m_{N_j}^2}\right),$$

where the symbols j and i label the decaying right-handed neutrino and the right-handed neutrinos that are exchanged within the loop, respectively. y_ν are the relevant Yukawa couplings and the function f is a factor that results from the integration over the loops [144]:

$$f(x) = x^{1/2}\left[\frac{2}{x-1} + \ln\left(\frac{1+x}{x}\right)\right]. \tag{12.32}$$

Focusing on the decays of the lightest of the three right-handed neutrinos, and working in the limit in which $m_{N_1} \ll m_{N_2}, m_{N_3}$, the loop function becomes $f \approx 3m_{N_1}/m_{N_i}$, and thus

$$\epsilon_{N_1} \approx -\frac{3}{8\pi} \sum_{i=2,3} \frac{\text{Im}[(y_\nu^\dagger y_\nu)_{i1}^2]}{(y_\nu^\dagger y_\nu)_{11}} \frac{m_{N_1}}{m_{N_i}}. \tag{12.33}$$

Relating the Yukawa couplings and the right-handed neutrino masses to the masses of the light neutrinos, we can use this expression to place an upper limit on the magnitude of ϵ_{N_1} [145]:

$$|\epsilon_{N_1}| \lesssim \frac{3}{8\pi} \frac{m_{N_1}}{v^2} (m_3 - m_1) \tag{12.34}$$

$$\lesssim 10^{-6} \times \left(\frac{m_{N_1}}{10^{10}\,\text{GeV}}\right)\left(\frac{m_3 - m_1}{0.05\,\text{eV}}\right),$$

where v is the vacuum expectation value of the Higgs field, and $m_3 - m_1$ is the splitting between the masses of the light neutrinos.

Once these decays have generated a lepton number asymmetry, that lepton number will be transformed through sphalerons into a mixture of net baryon and net lepton number, leading to a final effective value of $\epsilon = \epsilon_N \times (-28/51)$ (see Eq. 12.27). Combining this with the condition given in Eq. 12.34, we find that it is possible to generate enough net baryon number to achieve successful baryogenesis, $\epsilon \sim 2 \times 10^{-8}$, so long as the right-handed neutrinos in question are heavier than $m_N \gtrsim 2 \times 10^8$ GeV.

We thus conclude that the baryon asymmetry of our universe could have been generated through leptogenesis, relying on elements similar to those that are often invoked to explain the masses of the Standard Model neutrinos. More specifically, both require the existence of heavy right-handed neutrinos that are related to the light neutrinos through the seesaw mechanism. The main downside of traditional leptogenesis is that it does not make many concrete predictions for low-energy observables, making it difficult to test. Such considerations have motivated alternative versions of leptogenesis, which can involve right-handed neutrinos with masses near the TeV scale [146, 147, 148].

12.6 Electroweak baryogenesis

So far in this chapter, we've focused on scenarios in which Sakharov's third condition was satisfied by particles that decayed after falling out of equilibrium with the thermal bath. In this section, we will instead consider another class of models, in which the baryon asymmetry was generated during a departure from equilibrium that took place during the electroweak phase transition [149, 150, 151, 152, 153, 154].

In the context of quantum field theory, a phase transition is a process in which the universe (or a part of the universe) moves from one vacuum state into another. Second order phase transitions take place continuously, with the values of fields involved changing smoothly over time. In contrast, first order phase transitions are discontinuous, and rely on quantum tunneling. To satisfy Sakharov's third condition, a phase transition must have been strongly first order. In the Standard Model, the electroweak phase transition was *not* strongly first order, and therefore new physics is required to facilitate viable electroweak baryogenesis. Strongly first order phase transitions are of great interest to cosmologists not only for their potential role in baryogenesis, but also for the possibility that they could generate detectable gravitational wave signals, or leave behind strange relics known as topological defects.

Before moving on to consider electroweak baryogenesis itself, it will be helpful for us to first review some general aspects of phase transitions. Consider an example featuring a real scalar field, ϕ, that is described by the following Lagrangian:

$$\mathcal{L} = \frac{1}{2}\partial_\mu \phi \, \partial^\mu \phi - V(\phi), \tag{12.35}$$

where the low-temperature potential, V, takes a form that is given by

$$V(\phi) = V_0 - \frac{1}{2}m^2\phi^2 + \frac{1}{4}\lambda\phi^4. \tag{12.36}$$

Given enough time, ϕ will evolve to a value that minimizes the potential. In other words, the universe will tend toward a ground state. In some circumstances, however, the value

of ϕ can be temporarily displaced from the ground state, perhaps being trapped in a local minimum of the potential, or being in the process of rolling down a slope of the potential toward one of the minima. The potential described by Eq. 12.36 has three local extrema (identified by the condition $\partial V/\partial\phi = 0$), located at $\phi = 0$, $\sqrt{m^2/\lambda}$, and $-\sqrt{m^2/\lambda}$. Of these extrema, the latter two are minima ($\partial^2 V/\partial\phi^2 > 0$). Due to the structure of this potential, the value of ϕ will tend to evolve toward one or the other of these two ground states. Since the value of ϕ is not zero at these minima, the scalar field described by this theory can be said to have a finite vacuum expectation value, given by $v = \langle\phi\rangle = \pm\sqrt{m^2/\lambda}$.

The term proportional to ϕ^2 in Eq. 12.36 is responsible for the mass of the scalar described by this theory. More specifically, the square of the scalar's mass is equal to the second derivative of the potential, evaluated at either minimum:

$$\begin{aligned} m_\phi^2 &= \frac{\partial^2 V}{\partial\phi^2}(v) \\ &= -m^2 + 3\lambda\left(\frac{m^2}{\lambda}\right) \\ &= 2m^2. \end{aligned} \tag{12.37}$$

The rightmost term in Eq. 12.36 characterizes the interactions that take place between these scalars. In a more realistic theory, interactions involving other particles will also impact the shape of the potential.

In either of the minimum energy configurations, the energy density of the vacuum is given by $\rho_{\text{vac}} = V(\phi = |v|) = V_0 - m^4/4\lambda$. For the purposes of this discussion, I will set $V_0 = m^4/4\lambda$, so that the energy density of the vacuum is equal to zero in either ground state. With this choice made, we can now express the potential in terms of the vacuum expectation value, v, and the normalization, λ,

$$V(\phi) = \frac{\lambda}{4}(\phi^2 - v^2)^2. \tag{12.38}$$

Notice that although the Lagrangian of this theory (as given in Eq. 12.35) is perfectly symmetric with respect to operations of the form $\phi \leftrightarrow -\phi$, the universe it describes cannot be. More specifically, at any given point in space, ϕ will be in *one or the other* of the two ground states, thus breaking the symmetry between positive and negative values of ϕ. In a situation such as this, in which the Lagrangian of a theory has a symmetry that is not respected by the vacuum, one says that that symmetry has been spontaneously broken.

At high temperatures, interactions between particles can cause the shape of the potential to be modified [155]. Although a detailed discussion of finite-temperature field theory is well beyond the scope of this book, we can get a flavor for some of the dynamics involved by examining a couple of simple examples. In particular, let's consider a potential with the same low-temperature behavior as described above, but with the following finite-temperature corrections:

$$\begin{aligned} V_T(\phi, T) &= V(\phi) + \alpha\, T^2\phi^2 \\ &= \frac{\lambda v^4}{4} + \left(\alpha\, T^2 - \frac{\lambda v^2}{2}\right)\phi^2 + \frac{\lambda}{4}\phi^4, \end{aligned} \tag{12.39}$$

where $V(\phi)$ is the low-temperature potential. Evaluating this expression at $T = 0$, the thermal corrections vanish and we recover Eq. 12.38. This potential, which is plotted in the

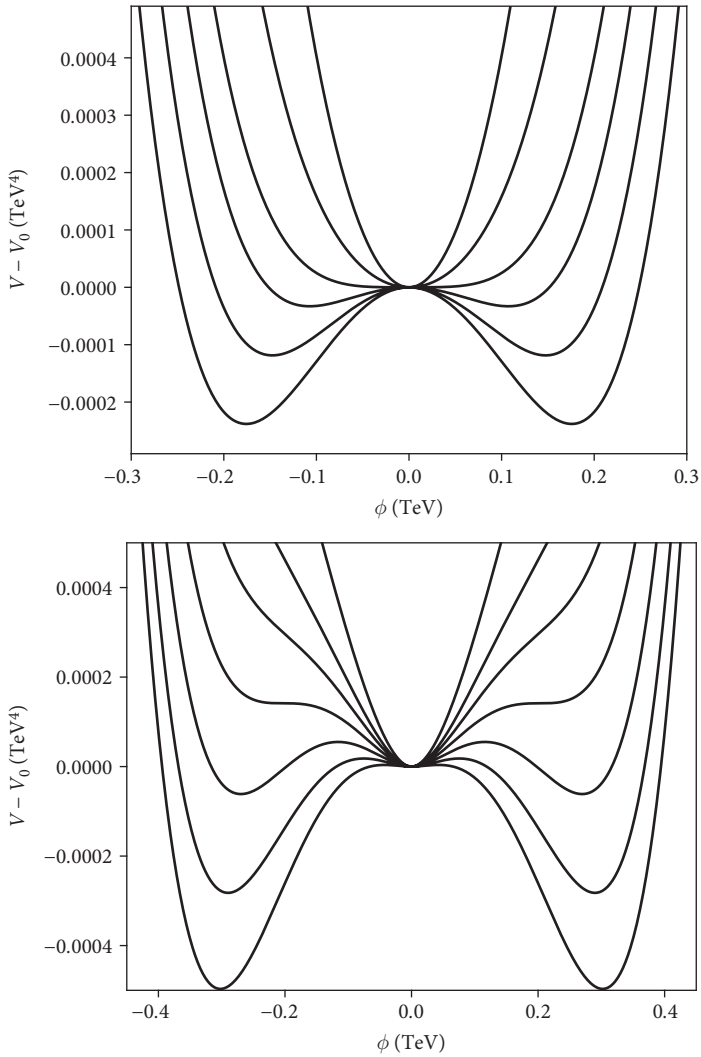


Figure 12.4 Examples of second order (top) and first order (bottom) phase transitions. In the top frame, the results correspond to the potential given in Eq. 12.39 with $\lambda = 1, \alpha = 0.25$, and $v = 246$ GeV. The curves from bottom to top in this frame are for $T = 0.7\,T_c$, $0.8\,T_c$, $0.9\,T_c$, $1.0\,T_c$, $1.2\,T_c$, and $1.5\,T_c$, where $T_c = \sqrt{\lambda v^2/2\alpha}$. In the bottom frame, the potential given in Eq. 12.40 is shown for $\lambda = 1$, $\alpha = 0.25$, $\beta = 0.3$, and $v = 246$ GeV. From bottom to top, these curves represent $T = 0.85\,T_1$, $T = 0.9\,T_1$, $0.95\,T_1$, $1.0\,T_1$, $1.05\,T_1$, $1.1\,T_1$, and $1.2\,T_1$, where $T_1 = 2\lambda v/\sqrt{8\alpha\lambda - 9\beta^2}$. In each frame, V_0 represents the value of the potential evaluated at $\phi = 0$.

upper frame of Figure 12.4, has local extrema at $\phi = 0$, and possibly at $\phi = \pm\sqrt{v^2 - 2\alpha T^2/\lambda}$. At temperatures above the critical temperature, $T > T_c = \sqrt{\lambda v^2/2\alpha}$, the only minimum is at $\phi = 0$, making that configuration the sole ground state. As the temperature drops below the critical temperature, however, the $\phi = 0$ state becomes a local maximum, while the other two extrema (those at $\phi = \pm\sqrt{v^2 - 2\alpha T^2/\lambda}$) become the energetically preferred ground

states. Also notice that when evaluated at $T = T_c$, the value of ϕ is equal to zero at all three of the extrema, since $\phi = \pm\sqrt{v^2 - 2\alpha T_c^2/\lambda} = 0$. This demonstrates that the ground state transitions continuously between $\phi = 0$ at high temperatures to $|\phi| = \sqrt{v^2 - 2\alpha T^2/\lambda} > 0$ at temperatures below T_c. The continuous nature of this change makes it an example of a second order phase transition.

Since we are interested here in phase transitions that are first order, let's modify the potential to include a new term that is proportional to $-T|\phi|^3$:

$$V_T(\phi, T) = \frac{\lambda v^4}{4} + \left(\alpha T^2 - \frac{\lambda v^2}{2}\right)\phi^2 - \beta T|\phi|^3 + \frac{\lambda}{4}\phi^4. \tag{12.40}$$

For this potential, the local extrema are found at $\phi = 0$, and at $\phi = \pm[3\beta T \pm \sqrt{(9\beta^2 - 8\lambda\alpha)\, T^2 + 4\lambda^2 v^2}\,]/2\lambda$. Notice that the later extrema are real only if the following condition is satisfied:

$$T < T_1 = \frac{2\lambda v}{\sqrt{8\alpha\lambda - 9\beta^2}}. \tag{12.41}$$

At temperatures above T_1, the potential's only extremum is the minimum found at $\phi = 0$. At lower temperatures, there are three distinct minima (and two maxima between them), one at $\phi = 0$ and the other two at $\phi = \pm(3\beta T/2\lambda + v\sqrt{1 - T^2/T_1^2})$. This transition is illustrated in the lower frame of Figure 12.4.

A critical point here is that, for this choice of potential, there can exist barriers that separate the high-temperature minimum (at $\phi = 0$) from what become the new global minima when $T = T_c$. These barriers can temporarily prevent ϕ from transitioning into either of the global minima. More specifically, if a $V > V_0$ barrier exists between the local minimum (at $\phi = 0$) and the global minima (at $|\phi| = [3\beta T + \sqrt{(9\beta^2 - 8\lambda\alpha)\, T^2 + 4\lambda^2 v^2}\,]/2\lambda$), the transition from the local minimum to either global minimum will not be able to occur classically. Instead, such a transition can only take place through quantum tunneling.

When ϕ tunnels into one of the global minima, a bubble forms at that point in space. Inside of that bubble, the value of the field is given by $|\phi| = 3\beta T/2\lambda + v\sqrt{1 - T^2/T_1^2}$, corresponding to the global minimum of the potential. In the space outside of the bubble, the vacuum is still in the local minimum, with $\phi = 0$. Because the global minima are energetically preferred, the bubble grows, expanding outward in all directions at the speed of light. The nucleation of these bubbles doesn't necessarily start as soon as an energetically favorable minimum forms, at $T = T_c$. Instead, this process only takes off when the tunneling rate becomes larger than the rate of Hubble expansion. In some situations, many bubbles can form and even collide with one another as they expand. This continues until the local minimum disappears, at which point the value of ϕ is able to move classically toward the global minima throughout all of space. It is through this process of bubble nucleation that a first order phase transition would proceed in the early universe, potentially providing us with conditions that could satisfy Sakharov's third condition [156, 157].

We're now in a position to discuss the specific case of the electroweak phase transition. As it turns out, the Higgs potential in the Standard Model in the temperature range of interest can be approximately expressed in the same form as Eq. 12.40 [158]:

$$V(h, T) \approx \frac{\lambda v^4}{4} + \left(\alpha T^2 - \frac{\lambda v^2}{2}\right)h^2 - \beta T|h|^3 + \frac{\lambda}{4}h^4, \tag{12.42}$$

where h is the value of the Higgs field. The relevant parameters of this potential can be related to those of the Standard Model's electroweak sector:

$$\lambda = \frac{m_h^2}{2v^2} \approx 0.13, \tag{12.43}$$

$$\alpha = \frac{g_1^2 + 3g_2^2 + 4y_t^2 + 8\lambda}{32} \sim 0.2,$$

$$\beta = \frac{3}{96\pi}\left(2g_1^3 + (g_1^2 + g_2^2)^{3/2}\right) \approx 0.0096,$$

where $y_t \sim 1$ is the top quark's Yukawa coupling and $m_h \approx 125\,\text{GeV}$ is the mass of the Higgs boson.

For these parameter values, this phase transition is only very weakly first order. This can be quantified in terms of the order parameter, which is equal to the ratio of the field value to the temperature, as evaluated at the critical temperature (i.e., the temperature at which the $h > 0$ minima become the global minimum). For the potential described by Eq. 12.42, the order parameter is $h_c/T_c \approx 20\,\text{GeV}/140\,\text{GeV} \sim 0.14$. In contrast, in order for the electroweak phase transition to provide a viable environment for baryogenesis, the order parameter must satisfy $h_c/T_c \gtrsim 1.09$ [159]. If this condition is not met, sphaleron transitions within the bubbles will be able to erase any baryon asymmetry that may have otherwise been generated.

The real problem with the Standard Model Higgs potential is that, at the critical temperature, there is only a very small barrier to prevent tunneling into the true vacuum. This occurs at $|h| \approx 20\,\text{GeV}$ and has a maximum height of only $V \approx (4\,\text{GeV})^4$. This is not nearly enough to keep the Higgs field in its false vacuum state and thus leads to a phase transition that is nearly continuous. In fact, more detailed calculations indicate that no barrier exists in this case, making this an example of a second order phase transition [154].

In order for the electroweak phase transition to have been strongly first order, we need to introduce new physics that changes the shape of the Higgs potential. More specifically, we can obtain a sufficiently strong first order phase transition if some kind of new physics increases the value of β by a factor of ~ 5 or more. For a value of $\beta = 0.1$, for example, we obtain an order parameter of $h_c/T_c \approx 260\,\text{GeV}/140\,\text{GeV} \sim 1.85$, which is large enough to achieve viable baryogenesis. Notice that the potential barrier at $T = T_c$ is much higher in this case, $V \approx (55\,\text{GeV})^4$, allowing it to adequately suppress tunneling into the true vacuum.

Historically, the most popular way of increasing the magnitude of the $-T|h|^3$ term in the Higgs potential has been to introduce light stops (the supersymmetric partners of the top quark; see Sec. 17.1) [160]. In order for the interactions of the stops to increase the value of β enough to facilitate electroweak baryogenesis, however, these particles must be lighter than the top quark. This has essentially been ruled out by searches for such particles at the Large Hadron Collider [161, 162]. In light of these results, modern models of electroweak baryogenesis instead feature relatively light particles that do not appreciably interact with the Standard Model, except through their couplings to the Higgs.

Up to this point, our discussion of electroweak baryogenesis has focused on Sakharov's third condition. Even in the presence of a sufficiently strong first order phase transition, the other two conditions for effective baryogenesis will also need to be satisfied. Fortunately, electroweak interactions do, in fact, violate the conservation of baryon number at high temperatures, automatically satisfying Sakharov's first condition. Furthermore, it seems

quite plausible that the interactions of the Higgs boson, or other scalars with masses near the electroweak scale, could violate the symmetry of CP. If this is the case, then the collisions of these particles at the bubble walls will generate an asymmetry in the number densities of particles and antiparticles in those regions of space. These asymmetries will then diffuse away from the walls, into the region filled with the $h = 0$ vacuum. In this vacuum state, electroweak sphalerons will be efficient, transforming any matter-antimatter asymmetry into an asymmetry between the numbers of baryons and antibaryons. As those baryons and antibaryons get swept up by the expanding bubble walls, they become immune to the effects of sphalerons, and the asymmetry gets frozen in.

12.7 Topological defects

Following a phase transition, the fields at any given point in space will evolve toward a configuration that corresponds to one of the low-temperature ground states of the vacuum. There is no guarantee, however, that the *same* minimum energy ground state will be selected throughout all of space. In some cases, this can lead to the formation of topologically stable configurations of gauge and Higgs fields, known as topological defects [163, 164, 165, 166, 167].

12.7.1 Domain walls

The simplest topological defects are domain walls, which can form when a discrete symmetry is spontaneously broken [165]. Consider, for example, the scalar field model described by the Lagrangian in Eq. 12.35 and repeated here:

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\,\partial^\mu\phi - V(\phi), \tag{12.44}$$

where the low-temperature potential takes the form

$$V(\phi) = V_0 - \frac{1}{2}m^2\phi^2 + \frac{1}{4}\lambda\phi^4. \tag{12.45}$$

Recall that this potential has two discrete low-temperature ground states, at $\phi = \sqrt{m^2/\lambda}$ and $-\sqrt{m^2/\lambda}$. After undergoing spontaneous symmetry breaking, it is possible that the universe could contain some regions in which $\phi = \sqrt{m^2/\lambda}$ and others with $\phi = -\sqrt{m^2/\lambda}$. Since the value of ϕ must vary continuously across space, the boundaries that separate these regions must include a surface along which $\phi = 0$. Such surfaces are called domain walls.

Consider a flat domain wall of infinite extent in the xy plane at $z = 0$. Infinitely far from the domain wall, the universe is in the ground state, with $\phi = \sqrt{m^2/\lambda}$ on one side, and $\phi = -\sqrt{m^2/\lambda}$ on the other. The equation of motion for the scalar field in this case is given by

$$\begin{aligned}\frac{\partial^2\phi}{\partial z^2} &= \frac{dV}{d\phi} \\ &= -m^2\phi + \lambda\phi^3,\end{aligned} \tag{12.46}$$

whose solution is as follows:

$$\phi(z) = \left(\frac{m^2}{\lambda}\right)^{1/2}\tanh\left(\frac{z|m|}{\sqrt{2}}\right). \tag{12.47}$$

Far away from the wall, the absolute value of the tanh function approaches unity, leading to $\phi \approx \pm\sqrt{m^2/\lambda}$. In the region near the wall, however, $|\phi|$ is much smaller, and thus the potential is much greater. This solution, therefore, corresponds to a thin surface with a large energy density that separates regions that are in different minima of the potential.

The total energy density of the scalar field is given by

$$\begin{aligned}\rho(z) &= \frac{1}{2}\left(\frac{d\phi}{dz}\right)^2 + V(\phi) \\ &= \frac{m^4}{2\lambda}\,\cosh^{-4}\left(\frac{z|m|}{\sqrt{2}}\right).\end{aligned} \tag{12.48}$$

Integrating this over all z, we can calculate the energy per surface area of the domain wall:

$$\begin{aligned}\eta &\equiv \int_{-\infty}^{\infty} \rho(z)\,dz \\ &= \frac{m^4}{2\lambda^2}\int_{-\infty}^{\infty} \cosh^{-4}\left(\frac{z|m|}{\sqrt{2}}\right) dz \\ &= \frac{2\sqrt{2}\,|m|^3}{3\lambda}.\end{aligned} \tag{12.49}$$

From this exercise, we conclude that domain walls should be expected to contain vast quantities of the energy. The energy of a single domain wall associated with the potential described above is equivalent to the following mass:

$$\begin{aligned}m_{\rm DW} &= \frac{2\sqrt{2}\,|m|^3 A}{3\lambda} \\ &\sim 10^{35}\,M_\odot \times \left(\frac{m}{\rm TeV}\right)^3 \left(\frac{0.1}{\lambda}\right)\left(\frac{A}{\rm Gpc^2}\right),\end{aligned} \tag{12.50}$$

where A is the surface area of the domain wall. Since this is many orders of magnitude larger than the total mass contained within our entire observable universe, it is clear that domain walls can exist only if they enclose very small regions of space, or are associated with a very low energy scale of symmetry breaking [168].

Since domain walls are expected to form whenever a discrete symmetry is spontaneously broken, the existence of such symmetries would appear to be problematic for cosmology [168]. In particular, after the spontaneous breaking of a discrete symmetry, we should expect the universe to contain many domain walls. Some of these walls will extend across the entire cosmological horizon, while others will take the form of bubbles that enclose regions that are in a particular minimum of the vacuum. The total energy density associated with these domain walls will evolve as the universe expands. Once the walls have become nonrelativistic, one generally expects this to include a factor of a^2 to account for the increasing surface area of the walls, as well as a factor of a^{-3} for their decreasing number density, for a total scaling of $\rho_{\rm DW} \propto a^{-1}$. In addition to these factors, the tension of a given domain wall can cause it to gradually straighten, further increasing this rate to $\rho_{\rm DW} \propto a^{-2}$ [169]. Since this rate is significantly slower than that experienced by either matter or radiation, even a small initial abundance of domain walls would quickly

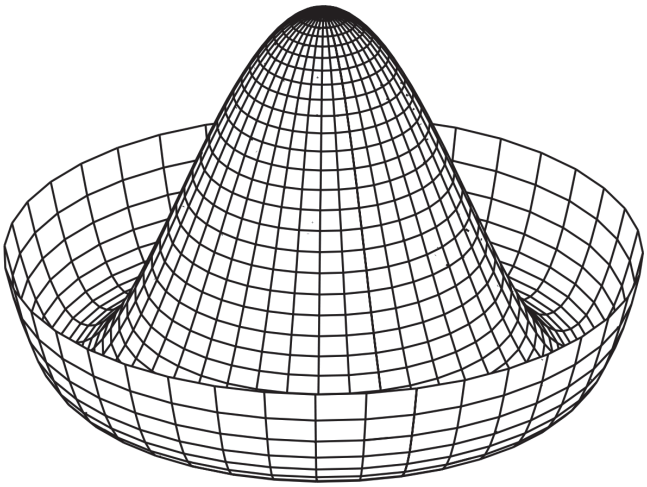


Figure 12.5 The "wine bottle" potential associated with a $U(1)$ symmetry, as described by Eq. 12.52. This potential has a continuous series of minima at $\Phi = m\, e^{i\theta}/(2\lambda)^{1/2}$ for all θ.

come to dominate the total energy density, allowing us to place strong constraints on any discrete symmetries that may have been spontaneously broken in the early universe. Such constraints could potentially be evaded, however, in models in which the domain walls are able to decay shortly after they form [170, 171, 172].

12.7.2 Cosmic strings

Domain walls are not the only kind of topological defects that are of interest to cosmologists. Whereas discrete symmetries can lead to the formation of two-dimensional domain walls, the spontaneous breaking of a $U(1)$ symmetry can result in one-dimensional defects known as cosmic strings [173]. Consider, for example, the following model consisting of a $U(1)$ gauge field, A_μ, and a complex Higgs field, Φ, which is charged under the $U(1)$:

$$\mathcal{L} = D_\mu \Phi D^\mu \Phi^\dagger - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} - V(\Phi), \tag{12.51}$$

where $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$, $D_\mu \Phi = \partial_\mu \Phi - ieA_\mu \Phi$, and e is the quantity of $U(1)$ charge that is carried by the Higgs field. The low-temperature potential in this model is given by

$$V(\Phi) = V_0 - m^2\, \Phi^\dagger \Phi + \lambda\, (\Phi^\dagger \Phi)^2. \tag{12.52}$$

Potentials of this form are sometimes described as having a "wine bottle" or "Mexican hat" shape (see Figure 12.5), and feature minima at $|\Phi| = m/(2\lambda)^{1/2}$. In our discussion of domain walls, the potential of the real scalar field had two discrete minima. In this case, the complex scalar field instead has an infinite number of continuously varying minima, corresponding to $\Phi = m\, e^{i\theta}/(2\lambda)^{1/2}$ for all values of θ between 0 and 2π. Throughout this entire collection of minima, this theory contains a Higgs boson with a mass given by $m_h = \sqrt{2}\, m$, and a vector boson with a mass of $m_V = e\, m/\sqrt{\lambda}$.

Although the value of θ can vary across space, it must do so continuously. This means that around any closed loop, the value of θ must change by an integer multiple of 2π. Such a loop can grow or shrink, but will always remain finite in length. To be clear, this loop is *not*

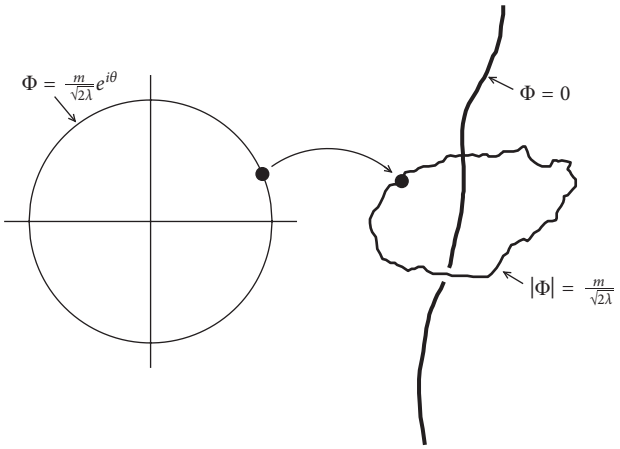


Figure 12.6 In theories with a spontaneously broken $U(1)$ symmetry, closed loops around the minimum of the potential (where $|\Phi| = m/\sqrt{2\lambda}$) necessarily enclose a curve along which $\Phi = 0$. These one-dimensional objects, along which the universe is in a state of false vacuum, are known as cosmic strings.

a cosmic string. After all, the field is at a minimum of the potential at every point in space along this loop. Notice, however, that *through* each such loop, there must pass a curve along which $\Phi = 0$. It is along *this* curve that the scalar field is not in one of its minima, and it is this structure that constitutes a cosmic string (see Figure 12.6).[2]

Cosmic strings generally contain much less energy than domain walls. Although the details of the relevant calculations are involved, they show that the energy per length of a cosmic string is expected to be on the order of $\mu \sim \mathcal{O}(\pi m^2/\lambda)$. Using this, we can estimate that a closed string loop of radius r should contain energy equivalent to the following mass:

$$m_{\rm string} \sim \frac{\pi m^2}{\lambda} 2\pi r \tag{12.53}$$
$$\approx 3 \times 10^{-8}\, M_\odot \times \left(\frac{m}{\rm TeV}\right)^2 \left(\frac{0.1}{\lambda}\right) \left(\frac{r}{\rm Gpc}\right).$$

Unlike domain walls, the existence of a modest number of cosmic strings in our universe would not necessarily be problematic for cosmology.

Similarly, to domain walls, any cosmic strings that were produced in the early universe will evolve as the universe expands [174, 175, 176]. While closed strings tend to become smaller with time, infinite strings (or strings that extend beyond the present horizon) tend to straighten. Numerical studies of these processes find that a cosmic string network will evolve to become a large collection of smaller closed strings, along with ~ 40 strings which extend across the observable universe [177, 178, 179, 180].

The presence of cosmic strings in our universe could have had a measurable impact on the CMB, in particular if the strings result from a symmetry that was spontaneously broken at a high energy scale. Before detailed measurements of the CMB's angular power spectrum were made, it was often speculated that cosmic strings might have played an important

[2] Cosmic strings are distinct from and not to be confused with the objects studied in string theory.

role in the process of structure formation. Modern measurements of the CMB, however, have ruled out this possibility, placing constraints on the order of $\mu/M_{\rm Pl}^2 \lesssim 10^{-7} - 10^{-6}$, corresponding to $m \lesssim 10^{16}\,{\rm GeV} \times \lambda^{1/2}$ [181]. At this upper limit, roughly 1% of the the CMB's angular power spectrum at $l = 10$ is the consequence of cosmic strings.

In addition to their impact on the CMB, oscillating loops of cosmic strings could generate potentially observable signals of gravitational waves [182, 183]. Furthermore, typical string loops contain features known as "kinks" and "cusps," which can lead to bursts of gravitational wave emission [184]. Observations of pulsars have been used to place constraints on this signal at a level of $\mu/M_{\rm Pl}^2 \lesssim 10^{-9}$, at least in some scenarios [177, 185]. Searches for short duration gravitational wave bursts by the Advanced LIGO-Virgo observatory have also been used to constrain cosmic strings. While these constraints are somewhat model-dependent, they can in some cases be as stringent as $\mu/M_{\rm Pl}^2 \lesssim 10^{-15}$ [186].

Lastly, if a cosmic string lies along or near our line of sight to a distant light source, the presence of that string can alter the trajectory of the passing light, leading to distinctive patterns of gravitational lensing [187, 188]. Next generation gravitational lensing surveys operating at radio frequencies are expected to be sensitive to cosmic strings at a level of around $\mu/M_{\rm Pl}^2 \lesssim 10^{-9}$ [189].

12.7.3 Monopoles

In addition to two-dimensional domain walls and one-dimensional cosmic strings, spontaneous symmetry breaking can lead to the formation of point-like topological defects, known as monopoles. This is the case, for example, in theories in which an $SU(2)$ or $SO(3)$ gauge symmetry is spontaneously broken down to $U(1)$ by a Higgs triplet, Φ. After symmetry breaking, such a theory will include two massive and one massless gauge boson, as well as a massive Higgs boson. While the magnitude of Φ is determined by the minimization of the potential, its orientation in $SU(2)$ or $SO(3)$ space is not. The total energy is minimized in the case in which Φ is constant over all space. There are, however, configurations of the Higgs field that cannot be continuously deformed into this state, leading to the existence of stable topological defects. As it turns out, these defects are point-like, and contain energy equivalent to a mass of $m_M \sim 4\pi\langle\Phi\rangle/e \sim 4\pi m_V/e^2$, where e is the charge associated with the resulting $U(1)$ symmetry, and m_V is the mass of the corresponding gauge bosons. If we identify this $U(1)$ with the Standard Model's $U(1)_{\rm EM}$, then e is the electric charge of the electron, and the monopoles carry a nonzero amount of magnetic charge, $Q_M = 4\pi/e$. In such a model, the point-like topological defects are known as magnetic monopoles.

Monopoles arise whenever a gauge group is broken to a subgroup that includes a $U(1)$ [190, 191, 192]. From this, it follows that essentially all grand unified theories predict the existence of magnetic monopoles. Furthermore, the mass of these monopoles can be related to the gauge coupling of the GUT force and the masses of the GUT bosons:

$$\begin{aligned} m_M &\sim \frac{4\pi M_{X,Y}}{g_{\rm GUT}^2} \\ &\approx 8 \times 10^{19}\,{\rm GeV} \times \left(\frac{0.04}{g_{\rm GUT}}\right)^2 \left(\frac{M_{X,Y}}{10^{16}\,{\rm GeV}}\right). \end{aligned} \tag{12.54}$$

As it turns out, this prediction poses a significant challenge for Big Bang cosmology. In particular, shortly after the breaking of the GUT symmetry, the universe should have contained roughly one monopole in each horizon-sized region [165]. Since $d_h \sim H^{-1} = [45 m_{\rm Pl}^2/4\pi^2 g_\star(T) T^4]^{1/2}$, the resulting number density of monopoles per unit entropy is

given by

$$\frac{n_M}{s} \sim \left(\frac{3}{4\pi\, d_h^3}\right)\left(\frac{45}{2\pi^2 g_{\star,S}(T)\, T^3}\right) \tag{12.55}$$
$$\sim 4\times 10^{-9} \times \left(\frac{T}{10^{16}\,\text{GeV}}\right)^3 \left(\frac{g_\star}{400}\right)^{1/2},$$

where T is the temperature at which GUT symmetry breaking takes place and the monopoles are formed. This corresponds to the following abundance of monopoles in the current universe:

$$\Omega_M = \frac{m_M n_M}{\rho_{\text{crit}}} \tag{12.56}$$
$$= \frac{m_M}{\rho_{\text{crit}}}\left(\frac{n_M}{s}\right)\left(\frac{2\pi^2 g_{\star,S}(T_{\text{CMB}})T_{\text{CMB}}^3}{45}\right)$$
$$\sim 10^{20} \times \left(\frac{m_M}{8\times 10^{19}\,\text{GeV}}\right)\left(\frac{n_M/s}{4\times 10^{-9}}\right).$$

Thus, without some kind of mechanism to deplete their abundance, we should expect our universe to contain a density of monopoles that is $\sim 10^{21}$ times greater than that of matter.

You might think that the monopole abundance would be depleted through their annihilations. A comparison of this annihilation rate to the Hubble rate, however, finds this effect to be negligible, at least for values of n_M/s in the range described above [193]. The most well-known solution to the monopole problem is inflation, which, as we will see in the following chapter, can dilute the abundance of monopoles to acceptable levels.

Additional reading

There are many good references on baryogenesis in its various forms. For baryogenesis through out-of-equilibrium decays, I recommend Kolb and Turner's *The Early Universe*, as well as Riotto's summer school lectures on *Theories of Baryogenesis* [138]. For more background on grand unified theories, I recommend the Particle Data Group's review of this subject [24], as well as the classic text by Mohapatra, *Unification and Supersymmetry*. For the topic of leptogenesis, nice reviews by have been written by Buchmuller, Di Bari, and Plumacher [194] and Chen [195]. For a more recent summary, I'd recommend Cline's TASI lectures [196]. For more information on topological defects, I again direct the reader to Kolb and Turner's *The Early Universe*, as well as to the Particle Data Group's review of magnetic monopoles [24].

Problems

12.1 Consider a heavy gauge boson, X, that was in thermal equilibrium with the Standard Model bath in the early universe, and that fell out of equilibrium when $T \sim 0.1\, M_X$. Assuming that these particles do not appreciably decay, at what temperature would

the X population begin to dominate the energy density of the universe? What if the X population had instead frozen out of equilibrium when $T \sim 10\, M_X$?

12.2 In Sec. 12.3, I used $\Gamma_X \sim g_{\rm GUT}^2 M_X/4\pi$ for the decay rate of the heavy bosons, which is predicted in the case of renormalizable couplings. If the couplings are instead non-renormalizable, dimensional arguments lead us to estimate $\Gamma_X \sim M_X^3/M_{\rm Pl}^2$. Compare these two decay rates for $M_X \sim 10^{16}\,{\rm GeV}$ and $g_{\rm GUT} \sim 0.04$. For the non-renormalizable case, estimate the temperature at which these particles will depart from chemical equilibrium.

12.3 While the X bosons are falling out of equilibrium, 2-to-2 scattering processes mediated by these bosons can also increase or decrease the net baryon number of the universe. Estimating the cross section for this class of processes to be $\sigma \sim g_{\rm GUT}^4/(4\pi)^2 \times T^2/M_X^4$, compare the corresponding scattering rate to the decay rate of X.

12.4 In Sec. 12.3, we considered an example in which a net baryon asymmetry was generated through the interference terms between tree-level and one-loop diagrams for X and Y decay. Roughly speaking, how would this result change if such interference was to only arise at the two-loop level? If we hope to obtain a value of $\epsilon_{X,Y}$ that is large enough to generate the observed baryon asymmetry, what is the highest order of the interference terms that we could potentially rely on?

12.5 Consider a species of heavy boson that dominates the energy density of the universe before decaying. Assuming that this boson decays entirely to Standard Model particles, estimate the resulting value of Y_B as a function of the particle's mass, M, the average quantity of baryon number that is produced per decay, ϵ, and the temperature of the universe immediately after the decays.

12.6 If we could perfectly observe every particle within a radius equal to the Hubble length, estimate how long we would have to wait to observe one proton decay through a sphaleron transition.

12.7 Consider a scenario in which an order 1 asymmetry in $B+L$; $n_{B+L} \sim n \sim T^3$, was generated at a temperature of $T = 10^{10}\,{\rm GeV}$, where n is the total number density of all particles. Use Eq. 12.18 to estimate the temperature at which sphaleron transitions will have reduced the net baryon asymmetry to the measured value. What if the initial temperature was instead 1 TeV?

12.8 In deriving Eq. 12.27, I assumed that there are three generations of fermions. How would this result change if there were more than three generations?

12.9 For the potential given by Eq. 12.39, identify the extrema and determine which of these are minima, as a function of temperature.

12.10 For the Higgs potential given in Eq. 12.42, find the value of β that leads to an order parameter of 1.08. Hint: A relatively easy way to do this is to evaluate the potential numerically.

12.11 Confirm the results given in Eqs. 12.48 and 12.49.

12.12 Consider a scalar field with a potential given by $V = V_0 + \lambda\,[1 - \cos(N\phi/v)]$, where λ and v are free parameters, N is an integer, and ϕ is bounded to lie between 0 and $2\pi\, v$. 1) Identify the minima of this potential. 2) Show that $\phi(z) = (4v/N)\,\tan^{-1}(e^{Bz})$ is a solution to the equation of motion, and find the required value of B (z is the distance from a domain wall). 3) Calculate the energy density per surface area of the domain walls.

12.13 Generalize Maxwell's equations to allow for the possibility of nonzero magnetic charge.

12.14 Consider a population of GUT monopoles with a mass of $m_M = 10^{18}$ GeV that are produced at $T = 10^{16}$ GeV with an abundance given by $n_M/s = 10^{-10}$. Write down the Boltzmann equation for the evolution of their number density, including the effects of Hubble expansion and annihilation. For an annihilation cross section given by $\sigma v = (16\pi^2 T m_M)^{-1}$, compare the size of the competing terms and comment on the implications.

13

Inflation

The observed flatness and uniformity of our universe suggest that it once underwent a period of very rapid expansion, known as inflation. In light of inflation's empirical successes, it has become a central part of the modern cosmological paradigm. In this chapter, I describe the main motivations for inflation, and discuss how the evolution of a scalar field could have led to a period of exponential expansion at very early times. I also examine some particular models of inflation, and discuss observations that could be used to potentially discriminate between such models.

13.1 Motivations for inflation

With the right initial conditions, the Big Bang theory can explain the observed properties of the CMB, light element abundances, large-scale structure, and the expansion history of our universe. Despite these remarkable empirical successes, the initial conditions that one must adopt in order to explain these observations are widely perceived to be a cause for concern. In particular, for what one might consider to be "typical" initial conditions, we should expect our universe to be far less homogeneous and far more curved than it is observed to be. For these and other reasons, most modern cosmologists have come to the conclusion that our universe likely underwent a period of very rapid and exponential expansion at very early times [197, 198].

There are three lines of argument that are most often cited to motivate an era of cosmological inflation. In the following subsections, I'll describe these arguments as they have traditionally been formulated.

13.1.1 The flatness problem

The first Friedman equation (including curvature) can be written in the following form (see Eq. 2.51):

$$\Omega_\kappa \equiv 1 - \Omega_{\rm tot} = \frac{-\kappa}{a^2 H^2}. \tag{13.1}$$

From this, we see that the degree to which a universe is curved changes as it expands, in proportion to $a^{-2}H^{-2}$. For a universe with $|\Omega_\kappa| \ll 1$ (such as our own), the degree of

curvature grows as follows during periods of matter and radiation domination:

$$|\Omega_\kappa| \propto \frac{1}{a^2\rho} \propto \begin{cases} a & \text{(matter domination)}, \\ a^2 & \text{(radiation domination)}. \end{cases} \tag{13.2}$$

In order to explain the empirical fact that $|\Omega_\kappa| \lesssim 0.02$ in our universe today (see Chapter 7), $|\Omega_\kappa|$ must have been less than $\sim 10^{-5}$ at the time of recombination, and less than $\sim 10^{-16}$ during BBN. As we extrapolate to even earlier times, the maximum degree of curvature that is compatible with the observed characteristics of the CMB becomes increasingly miniscule. While there is nothing that strictly rules out the possibility that our universe started out with almost exactly zero curvature, this has struck many cosmologists as being implausibly fine-tuned. This realization has provided considerable motivation for cosmological models that could potentially explain why our universe is so flat.

One possible way to address the flatness problem would be to posit the existence of an era in the early universe during which the degree of curvature became *smaller* as space expanded. Since $|\Omega_\kappa| \propto a^{-2}\,\rho^{-1} \propto a^{-2+3(1+w)} = a^{1+3w}$ (see Eq. 2.47), this would require the equation of state to have been $w < -1/3$ during that era. For example, in an era that is dominated by constant density vacuum energy, $w = -1$, the degree of curvature evolves as $|\Omega_\kappa| \propto a^{-2}$. If such a state persisted long enough in the early universe, this could have reduced the magnitude of curvature to a such a degree as to cause our universe to be approximately flat today, without relying on fine-tuned initial conditions.

13.1.2 The horizon problem

Between time 0 and t, the maximum distance from which anything can reach us is limited by the cosmological horizon. The size of this horizon evolves with the expansion of the universe as follows (see Eq. 2.58):

$$\begin{aligned} d_h &= \int_0^t \frac{dt'}{a(t')} \\ &= \int_0^a \frac{da'}{Ha'^2} \propto \begin{cases} a^{1/2} & \text{(matter domination)}, \\ a & \text{(radiation domination)}. \end{cases} \end{aligned} \tag{13.3}$$

During periods of matter or radiation domination, the cosmological horizon grows monotonically. This means that there are regions of space in our universe that are currently within our horizon, but that were causally disconnected from each other in the past. For the measured values of Ω_M, $\Omega_{\rm rad}$, and Ω_Λ, the current size of this horizon is approximately $d_h \approx 13.4$ Gpc. At the time of recombination, however, the horizon was much smaller than it is today, $d_h \approx 0.24$ Mpc. This means that when we measure the CMB, we are detecting light from a large number of regions that were causally disconnected from one another (see Figure 13.1). Without any way for these regions to have interacted with each other, it is hard to understand why the CMB is so uniform in temperature.

Much as with the flatness problem, one possible way to resolve the horizon problem would be for there to have been a period in the early universe during which the size of the comoving horizon became smaller as space expanded. Writing Eq. 13.3 for a universe with

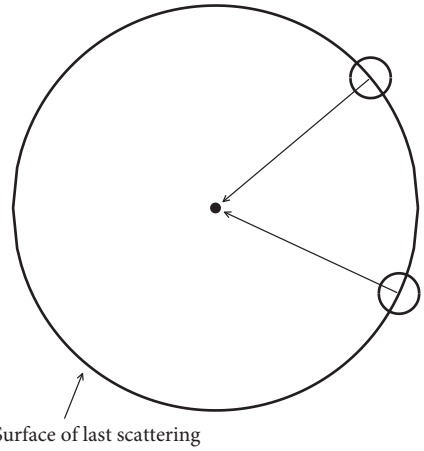


Figure 13.1 Despite the fact that the CMB is highly uniform, different locations along the surface of last scattering were well outside of each other's causal horizon at the time of recombination and decoupling.

a generic equation of state, w, we have

$$\begin{aligned} d_h &= \int_0^a \frac{da'}{Ha'^2} \\ &\propto \int_0^a \frac{da'}{\rho^{1/2}\, a'^2} \\ &\propto \int_0^a \frac{da'}{a'^{2-3(1+w)/2}} \\ &\propto a^{(1+3w)/2}. \end{aligned} \tag{13.4}$$

During periods with $w < -1/3$, the size of the comoving horizon shrinks as space expands. Thus, if such a state persisted for long enough, the entire volume of our observable universe could have been causally connected at very early times, providing a potential explanation for how our universe could be so homogeneous, despite having once consisted of many causally disconnected regions.

13.1.3 The problem of heavy relics

Many extensions of the Standard Model predict the existence of very heavy and stable states that would have been present at roughly equilibrium-level abundances in the very early universe. As we found in Sec. 9.5.2, if any of these particles are more massive than $\sim 10^2$ TeV, they would have survived the early universe with an abundance that exceeds the measured density of dark matter. Historically speaking, cosmologists in the late 1970s and 1980s were particularly concerned about magnetic monopoles in this context, which are predicted in many grand unified theories (see Sec. 12.7.3). In any case, if one thinks that heavy stable states are likely to exist, then it seems difficult to explain why we don't find large abundances of them in our universe today.

The fraction of the energy density that consists of a stable, decoupled, non-relativistic particle species, X, evolves as follows:

$$\frac{\rho_X}{\rho_{\rm tot}} \propto \frac{a^{-3}}{a^{-3(1+w)}} \tag{13.5}$$
$$= a^{3w} = \begin{cases} \text{constant} & \text{(matter domination)}, \\ a & \text{(radiation domination)}. \end{cases}$$

During periods that are dominated by matter or radiation, the relative abundance of these relics will either remain approximately constant, or increase as the universe expands. In contrast, the fraction of the energy density that consists of these relics will be reduced, or diluted, during any periods in which the equation of state is negative, $w < 0$. If the equation of state was negative after the creation of the relics in question, and remained negative for a sufficiently long period of time, this could potentially explain why no large abundances of heavy, stable relics are present in our universe today.

It should be noted that one possible solution to the problem of heavy relics is that perhaps no such heavy relics exist. If this were to turn out to be the case, then this motivation for inflation would simply be moot. In contrast, both the flatness and horizon problems really do seem to require some kind of new physics, such as that which could be provided by an era of inflation.

13.2 A common solution: accelerated expansion

In the previous section, we found that if there was a period in the early universe with $w < -1/3$, this could potentially explain why our universe is so flat and so uniform, as well as why it does not contain large abundances of exotic heavy relics. To help build our intuition for what a universe with $w < -1/3$ might be like, consider the second Friedmann equation:

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3M_{\rm Pl}^2}(\rho + 3p) \tag{13.6}$$
$$= -\frac{4\pi\rho}{3M_{\rm Pl}^2}(1 + 3w).$$

Written this way, it is clear that requiring $w < -1/3$ is the same as requiring $\ddot{a} > 0$. In other words, in order to solve the flatness and horizon problems, there must have been an era in the early universe in which the expansion rate was accelerating.

With the flatness, horizon, and heavy relics problems in mind, let's consider what kinds of physics might lead to a period of accelerating expansion. During an era with $w < -1/3$, the energy density of the universe decreases no more rapidly than $\rho \propto a^{-2}$, which is significantly slower than during periods of either matter or radiation domination. For example, in an era that is dominated by a substance with a constant energy density (such as that represented by a cosmological constant), the equation of state would be $w = -1$. Thus, to solve the flatness and horizon problems, there must have been a period in the early universe that was similar to this—dominated by something with a constant or slowly evolving energy density. Furthermore, this energy must have ultimately been transferred into the creation of a hot bath of matter and radiation, allowing for something like our universe to emerge.

During an era with an approximately constant energy density, $w \approx -1$, the first Friedmann equation yields the solution $a(t) \propto e^{t/\tau}$, where the timescale for exponential growth is determined by the energy density of the vacuum, $\tau = (3M_{\rm Pl}^2/8\pi\rho_\Lambda)^{1/2}$ (see Sec. 2.7.3). To see how a period of exponential expansion could potentially resolve the flatness, horizon, and heavy relics problems, let's posit that the period of vacuum domination persisted long enough for the scale factor to increase by N e-foldings, such that the duration of this era is equal to $\Delta t = N\tau$. During this time, the magnitude of curvature evolves as $|\Omega_\kappa| \propto a^{-2}$, reducing the degree of curvature by a factor of e^{2N}. Including this impact of inflation, we can calculate the degree of curvature that we should expect to find in our universe today:

$$\begin{aligned} |\Omega_\kappa| &\sim |\Omega_\kappa|_i \times e^{-2N} \times 10^{16} \left(\frac{T_{\rm rad}}{\rm MeV}\right)^2, \\ &\sim 10^{-10} \times |\Omega_\kappa|_i \left(\frac{e^{-2N}}{e^{-2\times 60}}\right)\left(\frac{T_{\rm rad}}{10^{10}\ {\rm GeV}}\right)^2, \end{aligned} \tag{13.7}$$

where $|\Omega_\kappa|_i$ is the initial magnitude of curvature and $T_{\rm rad}$ is the temperature of the radiation bath immediately after the period of exponential expansion. For the numbers I'm chosen in the second line ($N = 60$ e-foldings of exponential expansion and a subsequent temperature of $T_{\rm rad} = 10^{10}$ GeV), the magnitude of curvature is reduced by ten orders of magnitude over the course of the cosmic history. This would be more than enough to explain why our universe is indistinguishable from flat today, despite the growth in curvature that occurred during the radiation- and matter-dominated eras.

The horizon problem can be resolved by a period of inflation in a similar way. During a vacuum energy-dominated era, the size of the horizon evolves as $d_h \propto a^{-1}$, and thus *decreases* during that time by a factor of e^N. Including this, the size of the horizon today can be related to its initial value as follows:

$$\begin{aligned} d_h &\sim d_{h,i} \times e^{-N} \times \left(\frac{a_{\rm Eq}}{a_{\rm rad}}\right)\left(\frac{1}{a_{\rm Eq}}\right)^{1/2} \\ &\sim d_{h,i} \times e^{-N} \times \left(\frac{T_{\rm rad}}{T_{\rm Eq}}\right)\left(\frac{1}{3402}\right)^{1/2} \\ &\sim 10^{-5} \times d_{h,i} \left(\frac{e^{-N}}{e^{-60}}\right)\left(\frac{T_{\rm rad}}{10^{10}\ {\rm GeV}}\right), \end{aligned} \tag{13.8}$$

where $d_{h,i}$ is the size of the horizon at the start of exponential expansion. This illustrates that for a sufficiently long period of inflation (typically $N \sim 50$–60), the size of the horizon today will be *smaller* than it was initially. In other words, the entire observable universe was once in causal contact, resolving the horizon problem.

Lastly, we can evaluate the impact of inflation on the abundance of heavy relics. As an example, let's consider particles (or topological defects) with a mass near the scale of grand unification, $m_X \sim 10^{16}$ GeV. Let's also assume that these objects were present with roughly an equilibrium abundance as they were becoming nonrelativistic, $n_X \sim m_X^3$ at $T \sim m_X$. During N e-foldings of inflation, the abundance of these relics will be reduced by a factor of e^{3N}. After accounting for expansion during the radiation- and matter- dominated-eras, this

yields a current abundance of

$$\begin{aligned} n_X &\sim m_X^3 \times e^{-3N} \times \left(\frac{T_{\rm CMB}}{T_{\rm rad}}\right)^3 \\ &\sim 10^{-57}\,{\rm cm}^{-3} \times \left(\frac{m_X}{10^{16}\,{\rm GeV}}\right)^3 \left(\frac{e^{-3N}}{e^{-3\times 60}}\right) \left(\frac{10^{10}\,{\rm GeV}}{T_{\rm rad}}\right)^3 \\ &\sim 0.03\,{\rm pc}^{-3} \times \left(\frac{m_X}{10^{16}\,{\rm GeV}}\right)^3 \left(\frac{e^{-3N}}{e^{-3\times 60}}\right) \left(\frac{10^{10}\,{\rm GeV}}{T_{\rm rad}}\right)^3, \end{aligned} \tag{13.9}$$

corresponding to

$$\begin{aligned} \Omega_X &= \frac{m_X n_X}{\rho_{\rm crit}} \\ &\sim 2 \times 10^{-36} \times \left(\frac{m_X}{10^{16}\,{\rm GeV}}\right)^4 \left(\frac{e^{-3N}}{e^{-3\times 60}}\right) \left(\frac{10^{10}\,{\rm GeV}}{T_{\rm rad}}\right)^3. \end{aligned} \tag{13.10}$$

13.3 The dynamics of a slowly rolling scalar field

The fact that inflation did not last forever means that it must not have been driven by simple vacuum energy with a fixed and unchanging density. In order for a universe to transition out of an inflationary era, the energy responsible for this exponential expansion must somehow be transformed into a bath of energetic particles, leaving that universe in a hot, radiation-dominated state. If this phase transition were first order, it would tend to leave the universe in a state with large inhomogeneities, inconsistent with observations of our own universe [199]. A second order phase transition, in contrast, could allow a universe to smoothly transition out of its inflationary era, leading to predictions that could potentially be consistent with measurements of the CMB and large-scale structure [200, 201].

In order for inflation to solve the flatness and horizon problems, while still ultimately leaving our universe in a highly uniform, hot, and radiation-dominated state, one generally invokes a slowly evolving scalar field, known as the inflaton. During inflation, the energy density of the inflaton behaves much like a cosmological constant, remaining approximately constant and causing space to expand exponentially. As time passes, however, this scalar field gradually transitions from its initial state toward one with less potential energy, ultimately causing inflation to come to an end. Putting this in other words, the era of inflation is characterized by the inflaton slowly rolling down a nearly flat potential.

The energy density and pressure of the scalar field, ϕ, are given by

$$\begin{aligned} \rho_\phi &= \frac{1}{2}\dot{\phi}^2 + V(\phi), \\ p_\phi &= \frac{1}{2}\dot{\phi}^2 - V(\phi), \end{aligned} \tag{13.11}$$

where the dotted quantities represent time derivatives, $\dot{\phi} \equiv d\phi/dt$, and $V(\phi)$ is the density of the potential energy associated with the scalar field. We can combine these quantities to

obtain an expression for the equation of state:

$$w = \frac{p_\phi}{\rho_\phi} = \frac{\frac{1}{2}\dot{\phi}^2 - V(\phi)}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}. \tag{13.12}$$

Thus, in order for $w < -1/3$, the potential energy has to dominate over the kinetic energy.

The dynamics of the scalar field are described by the following classical equation of motion:

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0, \tag{13.13}$$

where $H = (8\pi\ (\dot{\phi}^2/2 + V)/3M_{\rm Pl}^2)^{1/2}$ is the rate of Hubble expansion.[1] During slow-roll inflation, the $\ddot{\phi}$ term in this expression is small, leading to $3H\dot{\phi} \approx -dV/d\phi$, and thus the $3H\dot{\phi}$ term acts as friction, slowing the rate at which the value of ϕ evolves.

Using the second Friedmann equation, we can write

$$\begin{aligned} \frac{\ddot{a}}{a} &= -\frac{4\pi}{3M_{\rm Pl}^2}(\rho + 3p) \\ &= -H^2\left(\frac{1+3w}{2}\right) \\ &= H^2\,(1-\epsilon), \end{aligned} \tag{13.14}$$

where

$$\begin{aligned} \epsilon &\equiv \frac{3}{2}\,(w+1) \\ &= \frac{3}{2}\left(\frac{\dot{\phi}^2}{\frac{1}{2}\dot{\phi}^2 + V(\phi)}\right) \\ &= \frac{4\pi\dot{\phi}^2}{H^2 M_{\rm Pl}^2}. \end{aligned} \tag{13.15}$$

The quantity ϵ is one of two slow-roll parameters that are traditionally used to characterize the evolution of the inflaton. We can also use the fact that $\dot{H} = -4\pi\dot{\phi}^2/M_{\rm Pl}^2$ (see Problem 13.6) to write ϵ as follows:

$$\epsilon = -\frac{\dot{H}}{H^2}. \tag{13.16}$$

Written this way, we can see that ϵ is a measure of how much the Hubble rate is changing with respect to time. Furthermore, from the expression given above for $\dot{H}$, it follows that

[1] I warn the reader that while this book uses the system of natural units (in which $c = \hbar = 1$), many discussions of inflation instead employ reduced Planck units, which follow from the choice of defining $c = \hbar = M_{\rm Pl}^{\rm red} = 1$, where $M_{\rm Pl}^{\rm red} \equiv M_{\rm Pl}\,(\hbar c/8\pi)^{1/2} = M_{\rm Pl}/(8\pi)^{1/2}$ is the reduced Planck mass. There are, therefore, a number of expressions appearing in this discussion that differ from those found elsewhere in the literature by factors of $M_{\rm Pl}^2/8\pi$ (or, equivalently, by factors of $1/8\pi G$).

$dH/d\phi = \dot{H}/\dot{\phi} = -4\pi\dot{\phi}/M_{\rm Pl}^2$. We can use this to express ϵ in the following terms:

$$\begin{aligned}\epsilon &= \frac{4\pi\dot{\phi}^2}{H^2 M_{\rm Pl}^2} \\ &= \frac{4\pi}{H^2 M_{\rm Pl}^2}\left(-\frac{dH}{d\phi}\frac{M_{\rm Pl}^2}{4\pi}\right)^2 \\ &= \frac{M_{\rm Pl}^2}{4\pi}\left(\frac{1}{H}\frac{dH}{d\phi}\right)^2.\end{aligned} \tag{13.17}$$

Alternatively, we can express ϵ as the rate of change in the logarithm of the Hubble rate per e-folding:

$$\begin{aligned}\epsilon &= -\frac{\dot{H}}{H^2} \\ &\approx -\frac{1}{H}\frac{dH}{dN} \\ &= -\frac{d\ln H}{dN},\end{aligned} \tag{13.18}$$

where N is the number of e-foldings. Note that in the second step, I have used the fact that $\epsilon \ll 1$ during slow-roll inflation, allowing us to write $dN = H\,(1-\epsilon)^{1/2}dt \approx Hdt$ (see Problem 13.7).

Along with ϵ, it is useful to introduce a second slow-roll parameter, η, which is related to the rate of change of $dH/d\phi$:

$$\begin{aligned}\eta &\equiv \frac{M_{\rm Pl}^2}{4\pi H}\frac{d^2H}{d\phi^2} \\ &= -\frac{d\ln|dH/d\phi|}{dN}.\end{aligned} \tag{13.19}$$

In order for inflation to persist for $N = 50-60$ e-foldings or longer, the Hubble rate and its derivatives must each change only slowly. Slow-roll inflation thus corresponds to the case in which $\epsilon \ll 1$ and $\eta \ll 1$.

Alternatively, there is a set of slow-roll parameters that can be written purely in terms of the shape of the potential:

$$\begin{aligned}\epsilon_v &\equiv \frac{M_{\rm Pl}^2}{16\pi}\left(\frac{1}{V}\frac{dV}{d\phi}\right)^2 \\ \eta_v &\equiv \frac{M_{\rm Pl}^2}{8\pi V}\frac{d^2V}{d\phi^2}.\end{aligned} \tag{13.20}$$

In the slow-roll regime, these parameters are closely related to ϵ and η:

$$\begin{aligned}\epsilon &\approx \epsilon_v; \\ \eta &\approx \eta_v - \epsilon_v.\end{aligned} \tag{13.21}$$

These quantities are directly related to ϵ and η because it is the shape of the potential that dictates how ϕ evolves, and thus how H changes over time. In order for ϕ and H to change slowly, the potential has to be approximately flat. Note that since the inflaton's mass is related to the shape of the potential, $m_\phi^2 = d^2V/d\phi^2$, the requirement of $\eta_v \ll 1$ is equivalent to requiring that the inflation be much lighter than the magnitude of the Hubble parameter during inflation, $m_\phi \ll H$.

We are now in a position to evaluate how long an era of inflation will last. The number of e-foldings that occur between a time $t_{\rm CMB}$ and the end of inflation, $t_{\rm end}$, is given by

$$\begin{aligned} N &= \int_{t_{\rm CMB}}^{t_{\rm end}} H dt \\ &\approx -\int_{\phi_{\rm CMB}}^{\phi_{\rm end}} \frac{H}{\dot{\phi}}\, d\phi \\ &= -\int_{\phi_{\rm CMB}}^{\phi_{\rm end}} \left(\frac{4\pi}{\epsilon M_{\rm Pl}^2}\right)^{1/2} d\phi \\ &\sim \left(\frac{4\pi}{\epsilon M_{\rm Pl}^2}\right)^{1/2} \Delta\phi, \end{aligned} \tag{13.22}$$

where $\Delta\phi = \phi_{\rm CMB} - \phi_{\rm end}$ is the change in the value of ϕ over the course of inflation. From this, we see that in order to obtain the $N \gtrsim 50$–60 e-foldings that are required to solve the flatness and horizon problems, either ϵ must be very small ($\lesssim 0.01$), or $\Delta\phi$ must be larger than the Planck scale (constituting a transplanckian field excursion).

Using Eq. 13.20, we can also relate the number of e-foldings to the shape of the inflaton potential:

$$\begin{aligned} N &\approx -\int_{\phi_{\rm CMB}}^{\phi_{\rm end}} \left(\frac{4\pi}{\epsilon M_{\rm Pl}^2}\right)^{1/2} d\phi \\ &\approx -\int_{\phi_{\rm CMB}}^{\phi_{\rm end}} \frac{8\pi V}{M_{\rm Pl}^2} \left(\frac{dV}{d\phi}\right)^{-1} d\phi. \end{aligned} \tag{13.23}$$

One should keep in mind that $\phi_{\rm CMB}$ does not necessarily correspond to the value of the inflaton field at the beginning of inflation. Rather, it represents the value of ϕ evaluated at the time that the primordial density perturbations we observe in the CMB were generated.

13.4 Primordial perturbations from inflation

If we were to treat inflation as a purely classical phenomenon, then the value of the inflaton field could, at a particular point in time, potentially be the same at all locations throughout space. Taking into account quantum fluctuations, however, there will inevitably be spatial variations in the value of ϕ, along with corresponding variations in the energy density and spatial curvature. In most models of inflation, these quantum fluctuations lead to an approximately scale-invariant spectrum of primordial density perturbations, consistent with those observed in the CMB [202, 203, 204, 205, 206, 207]. This is one of the most impressive empirical successes of inflation. Furthermore, since the primordial perturbations on a given scale only start evolving when that mode enters the horizon, inflation correctly

predicts that there should exist superhorizon density perturbations that are in phase (see Sec. 7.4) [36]. No local subhorizon physics can account for this observation. It is worth emphasizing that these predictions were made prior to any observations of the CMB's temperature anisotropies. Inflation is thus not merely a theory that was designed or selected to explain existing observations, but rather one that made a series of concrete predictions that were later shown to be correct.

In the slow-roll regime, inflation leads to a spectrum of primordial density perturbations that is given by

$$\mathcal{P}_{\mathcal{R}}(k) \approx \frac{8V}{3\epsilon_v M_{\rm Pl}^4}, \tag{13.24}$$

where V and ϵ_v are each evaluated at the value of ϕ that corresponds to when the mode in question exited the horizon, $k = aH$. Following Eq. 7.18, we can also express this spectrum in terms of a power-law:

$$\mathcal{P}_{\mathcal{R}}(k) = A_s \left(\frac{k}{k_0}\right)^{n_s(k)-1}, \tag{13.25}$$

where

$$n_s(k) = n_s + \frac{1}{2}\frac{dn_s}{d\ln k}\ln\left(\frac{k}{k_0}\right) + \cdots, \tag{13.26}$$

and where n_s and $dn_s/d\ln k$ each denote the quantities as evaluated at $k = k_0$. Note that the normalization of the primordial power spectrum is well constrained by measurements of the CMB, $A_s = (2.105 \pm 0.030) \times 10^{-9}$ [8].

In the extreme slow-roll limit, inflation leads to a perfectly scale-invariant spectrum of primordial density perturbations, $n_s = 1$. The evolution of the inflaton field in most models, however, leads to slight departures from this prediction, generally tilting this spectral index toward smaller values, and thus featuring greater density fluctuations on larger scales. Quantitatively, we can calculate the value of n_s from the slow-roll parameters:

$$n_s \approx 1 - 6\epsilon_v + 2\eta_v. \tag{13.27}$$

One should keep in mind that the spectral index, n_s, can vary with scale, k. To calculate the value of n_s on a particular scale, one should find the value of ϕ at the time that modes of that scale exited the horizon, and use that value of ϕ to evaluate ϵ_v and η_v.

We can also use the slow-roll parameters to calculate the running of this spectral index:

$$\frac{dn_s}{d\ln k} \approx -8\epsilon_v\,(3\epsilon_v - 2\eta_v) - 2\xi^2, \tag{13.28}$$

where

$$\xi^2 = \frac{M_{\rm Pl}^4}{64\pi^2}\,\frac{1}{V^2}\frac{dV}{d\phi}\,\frac{d^3V}{d\phi^3}. \tag{13.29}$$

Measurements of the CMB's angular power spectrum by the Planck Collaboration have found $n_s = 0.9649 \pm 0.0042$, and $dn_s/d\ln k = -0.0045 \pm 0.0067$ [208]. This favors values of

$3\epsilon_v - \eta_v \sim 0.01$–$0.02$, at least during the period of inflation in which the primordial density perturbations observed in the CMB were generated.

13.5 Gravitational waves and primordial *B*-modes

In addition to scalar density perturbations, tensor perturbations were also generated during inflation, representing propagating gravitational waves. These gravitational waves can, in some inflationary models, leave observable levels of *B*-mode polarization in the CMB's angular power spectrum (see Sec. 7.7).

The magnitude of the primordial *B*-mode polarization signal can be quantified in terms of the tensor-to-scalar ratio, which is related to the value of the slow-roll parameter, ϵ :

$$r \equiv \frac{\mathcal{P}_T}{\mathcal{P}_\mathcal{R}} \approx 16\epsilon, \tag{13.30}$$

where $\mathcal{P}_T$ and $\mathcal{P}_\mathcal{R}$ are the power spectra for tensor and scalar modes, respectively. After fixing the amplitude of the scalar modes in order to match the magnitude of the observed temperature fluctuations in the CMB, r can be thought of as a measure of the tensor perturbations that are predicted by a given inflationary model. More specifically, the value of r is directly proportional to the energy density of the inflaton field during inflation, $\rho \approx V$:

$$r \sim 0.01 \times \frac{V}{(10^{16}\,\mathrm{GeV})^4}. \tag{13.31}$$

Alternatively, one can think of r as being proportional to the value of H^2 during inflation.

At present, the strongest constraint on r has been derived from a combination of data from BICEP/Keck and Planck, which finds $r \lesssim 0.032$ [209] (see also, Ref. [208]). As we will see in the section below, the tensor-to-scalar ratio provides us with a powerful way to distinguish between different classes of inflationary models.

13.6 Examples of inflationary models

Like dark matter candidates, numerous models of inflation have been proposed. In this section, I'll describe two such examples, which are intended to be representative of the broader classes of large-field and small-field inflation models.

13.6.1 Large-field inflation

Let's begin with the following example of a scalar field potential that leads to slow-roll inflation:

$$V(\phi) = \frac{1}{2} m_\phi^2 \phi^2. \tag{13.32}$$

By taking the second derivative of this potential, we confirm that the parameter m_ϕ is the mass of the inflaton. From this potential, we can use Eq. 13.20 to calculate the slow-roll

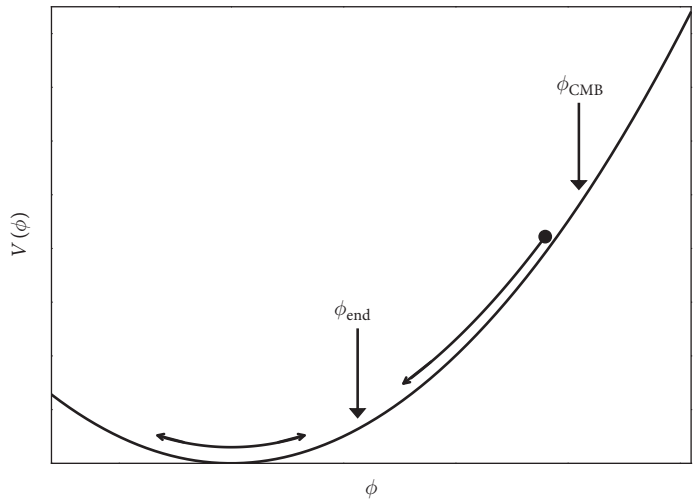


Figure 13.2 Inflation in a model with a potential given by $V(\phi) \propto \phi^2$. Between $\phi_{\rm CMB}$ and $\phi_{\rm end}$, 60 e-foldings of inflation take place. Once inflation ends, the field oscillates around the minimum of the potential, causing the energy density of the scalar field to be converted into a hot bath of radiation.

parameters:

$$\epsilon_v = \frac{M_{\rm Pl}^2}{4\pi\phi^2}, \tag{13.33}$$

$$\eta_v = \frac{M_{\rm Pl}^2}{4\pi\phi^2}.$$

From this, we see that in order to satisfy the slow-roll conditions ($\epsilon_v, \eta_v < 1$), the value of ϕ must be near, or larger than, the Planck scale.

The number of e-foldings of inflation that take place in this model between $\phi_{\rm CMB}$ and $\phi_{\rm end}$ is given by (see Eq. 13.22)

$$\begin{aligned} N_{\rm CMB} &\approx -\int_{\phi_{\rm CMB}}^{\phi_{\rm end}} \left(\frac{4\pi}{\epsilon M_{\rm Pl}^2}\right)^{1/2} d\phi \qquad (13.34)\\ &= -\int_{\phi_{\rm CMB}}^{\phi_{\rm end}} \frac{4\pi\phi}{M_{\rm Pl}^2}\, d\phi \\ &= \frac{2\pi}{M_{\rm Pl}^2}\,(\phi_{\rm CMB}^2 - \phi_{\rm end}^2) \\ &= \frac{2\pi\phi_{\rm CMB}^2}{M_{\rm Pl}^2} - \frac{1}{2}, \end{aligned}$$

where in the last step I have used the condition $\epsilon_v = 1$ to determine $\phi_{\rm end} = M_{\rm Pl}/2\sqrt{\pi}$. From this, we see that in order for $N_{\rm CMB} \sim 50$–60, we must require $\phi_{\rm CMB} \approx (N_{\rm CMB}/2\pi)^{1/2}\, M_{\rm Pl} \sim 2.8$–$3.1\, M_{\rm Pl}$ (see Figure 13.2).

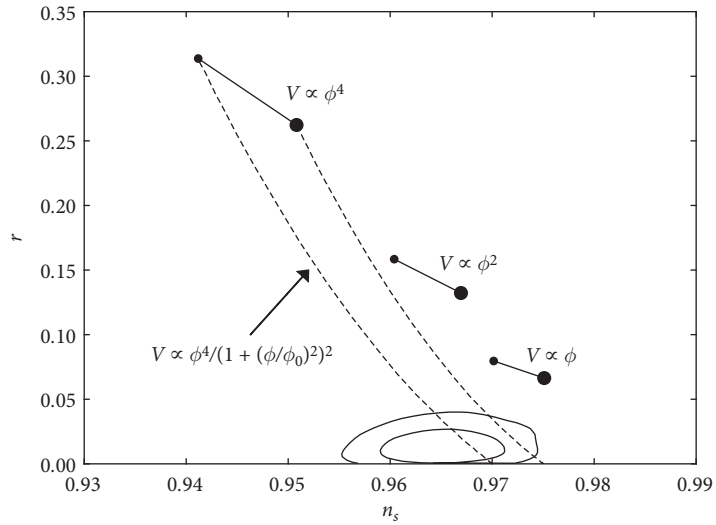


Figure 13.3 The spectral index of the primordial density perturbations and the tensor-to-scalar ratio predicted in selected models of inflation for $N_{\rm CMB} = 60$ (right) and $N_{\rm CMB} = 50$ (left). The measured values of these parameters are indicated by the closed contours [209].

In this model, we can use Eqs. 13.27 and 13.28 to determine the spectral index and the tilt of the primordial density perturbations:

$$n_s \approx 1 - 6\epsilon_v + 2\eta_v \tag{13.35}$$
$$= 1 - \frac{M_{\rm Pl}^2}{\pi\phi^2},$$

and

$$\frac{dn_s}{d\ln k} \approx -8\epsilon_v(3\epsilon_v - 2\eta_v) - 2\xi^2 \tag{13.36}$$
$$= -\frac{M_{\rm Pl}^4}{2\pi^2\phi^4}.$$

To compare these expressions with the values determined from the CMB, we evaluate the spectral index and tilt at $\phi = \phi_{\rm CMB}$, as calculated above. For this value of $\phi_{\rm CMB}$, we obtain $n_s \approx 1 - 2/N_{\rm CMB} \approx 0.960$–$0.967$ and $dn_s/d\ln k \approx -2/N_{\rm CMB}^2 \approx -(0.0006$–$0.0008)$, each of which are consistent with the measurements reported by the Planck Collaboration.

The problem with this model turns out to be that it predicts a large amplitude of tensor modes. The tensor-to-scalar ratio in this model is given by

$$r \approx 16\epsilon_v \tag{13.37}$$
$$= \frac{4M_{\rm Pl}^2}{\pi\phi^2}.$$

When evaluated at $\phi = \phi_{\rm CMB}$, this expression yields $r \approx 8/N_{\rm CMB} \approx 0.13$–$0.16$, well above the upper limits established by BICEP/Keck and Planck.

Similar conclusions are reached for many inflationary models that feature a polynomial potential [210]. In Figure 13.3, I have plotted the values of n_s and r that are predicted in models with $V \propto \phi^4$, $V \propto \phi^2$, and $V \propto \phi$. These models of inflation are each ruled out by the upper limits on r.

13.6.2 Small-field inflation

The models considered in the previous subsection ultimately failed to provide an acceptable description of inflation, as they predicted unacceptably large values for the tensor-to-scalar ratio, r. Putting this another way, inflation in those models generated the observed primordial density perturbations when the potential energy density was too large. To see why this is the case, we can use Eqs. 13.24 and 13.25 to write

$$\begin{aligned} V &\approx \frac{3}{8}\epsilon_v A_s M_{\rm Pl}^4 \\ &\approx (1.4 \times 10^{16}\,{\rm GeV})^4 \times \left(\frac{r}{0.032}\right), \end{aligned} \tag{13.38}$$

where in the second step I used the measured value of $A_s \approx 2.1 \times 10^{-9}$ and the relationship $r \approx 16\epsilon_v$. Thus, in order for an inflationary model to predict an acceptable tensor-to-scalar ratio, it must feature a relatively small potential energy density, $V(\phi_{\rm CMB}) \lesssim (10^{16}\,{\rm GeV})^4$. With this requirement in mind, it is common to classify models of inflation as either small-field or large-field models.

To be consistent with the upper limits on the tensor-to-scalar ratio, it is common to consider inflationary potentials that feature an approximately flat plateau, which then steepens at smaller values of ϕ, allowing inflation to come to an end. One such example is characterized by the following potential:

$$V(\phi) = \frac{V_0 \phi^4}{\left[1 + (\phi/\phi_0)^2\right]^2}, \tag{13.39}$$

where V_0 and ϕ_0 are free parameters of the theory. Taking the first and second derivatives of this potential, we can calculate the slow-roll parameters:

$$\begin{aligned} \epsilon_v &= \frac{M_{\rm Pl}^2}{\pi}\,\frac{\phi_0^4}{\phi^2(\phi^2+\phi_0^2)^2}, \\ \eta_v &= \frac{M_{\rm Pl}^2}{2\pi}\left[\frac{3}{\phi^2} - \frac{9}{\phi_0^2[1+(\phi/\phi_0)^2]} + \frac{6\phi^2}{\phi_0^4[1+(\phi/\phi_0)^2]^2}\right]. \end{aligned} \tag{13.40}$$

As we did in the case of $m_\phi^2\phi^2$ inflation, we can apply the criteria $\epsilon_v = 1$ to find the value of $\phi_{\rm end}$, and then use the requirement of $N_{\rm CMB} = 50 - 60$ to determine $\phi_{\rm CMB}$. In Figure 13.4, we show this potential and the points along it that correspond to $\phi_{\rm end}$ and $\phi_{\rm CMB}$, for the choice of $\phi_0 = 0.5 M_{\rm Pl}$.

The values of n_s, $dn_s/d\ln k$, and r that are predicted in this model depend on the choice of the parameter, ϕ_0, which in turn determines the steepness of the potential near $\phi = \phi_{\rm CMB}$. In the limit of large ϕ_0, the potential behaves as $V \propto \phi^4$, resulting in an unacceptably large value for the tensor-to-scalar ratio. For smaller values of ϕ_0, the potential is much flatter,

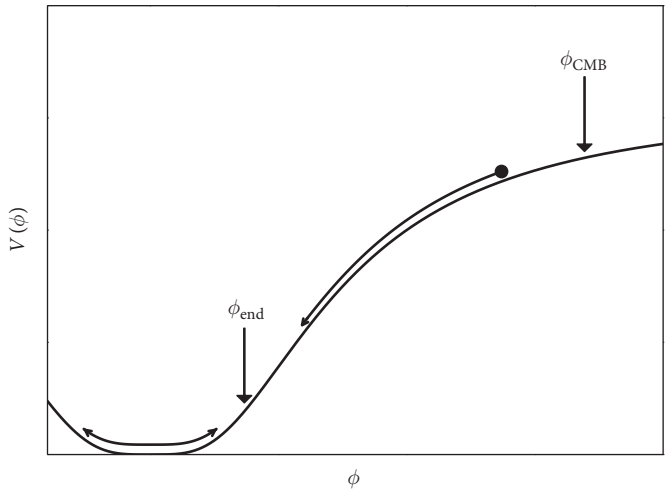


Figure 13.4 An example of a small-field inflation model with a potential given by $V \propto \phi^4/(1+(\phi/\phi_0)^2)^2$, shown for the case of $\phi_0 = 0.5\,M_{\rm Pl}$. This model features an approximately flat, plateau-like potential, followed by the onset of a steeper slope which causes inflation to end. Between $\phi_{\rm CMB}$ and $\phi_{\rm end}$, 60 e-foldings of inflation take place.

and thus the amplitude of the tensor modes is reduced, leading to consistency with current limits. For the case of $\phi_0 = 0.5 M_{\rm Pl}$, for example, we find $\phi_{\rm end} \approx 0.37 M_{\rm Pl}$, and $\phi_{\rm CMB} \approx 1.69 M_{\rm Pl}$ (using $N_{\rm CMB} = 60$). Using this value of $\phi_{\rm CMB}$ to evaluate the slow-roll parameters, we find $r \approx 0.015$, which is consistent with the current upper limits.

The values of n_s and r that are predicted in this model of inflation are shown in Figure 13.3. In the large ϕ_0 ($V \propto \phi^4$) limit, the predicted tensor-to-scalar ratio is unacceptably large, $r \sim 0.3$. If ϕ_0 is chosen to be smaller, however, the observational constraints on the tensor-to-scalar ratio can easily be accommodated.

13.7 Reheating

So far in this chapter, I have only mentioned in passing the transition that took place between the inflationary era and the subsequent reheating of the universe through the creation of a bath of hot particles. This transformation occurred once the inflaton field had rolled to the bottom of its potential and begun to oscillate about that minimum [136, 211]. For the sake of concreteness, let's consider a potential that can approximated by $V \propto \phi^2$ near its minimum. Inserting this into the equation of motion given in Eq. 13.13, we find that the inflaton evolves according to the following expression during the period of reheating:

$$\ddot{\phi} + 3H\dot{\phi} + \omega_0^2 \phi = 0, \tag{13.41}$$

where $\omega_0^2 = (dV/d\phi)/\phi$, which for this choice of potential is a constant. So long as $H < 2\omega_0/3$, the solution to this equation of motion is that of an underdamped harmonic oscillator:

$$\phi = \phi_0\, e^{-3Ht/2} \cos(\omega_1 t + \alpha), \tag{13.42}$$

where $\omega_1^2 = \omega_0^2 - (9/4)H^2 \approx \omega_0^2$. Furthermore, for $H \ll \omega_0$, the behavior of ϕ is approximately sinusoidal, $\phi \propto \cos(\omega_0 t + \alpha)$. By taking the time derivative of ϕ, we can use Eq. 13.11 to find the energy density during this era:

$$\rho_\phi \approx \frac{1}{2}\dot{\phi}^2 \approx \frac{1}{2}\left[-\phi_0 e^{-3Ht/2}\omega_0 \sin(\omega_0 t + \alpha) - \frac{3\phi_0 H}{2} e^{-3Ht/2}\cos(\omega_0 t + \alpha)\right]^2$$
$$\approx \frac{\phi_0^2 \omega_0^2}{2} e^{-3Ht}\sin^2(\omega_0 t + \alpha). \tag{13.43}$$

After averaging over one or more complete oscillations, $\langle \sin^2 x \rangle = 1/2$, we obtain the following expressions for the average energy density and its time derivative:

$$\bar{\rho}_\phi \approx \frac{\phi_0^2 \omega_0^2}{4} e^{-3Ht}, \tag{13.44}$$
$$\frac{d\bar{\rho}_\phi}{dt} \approx \frac{-3H\phi_0^2\omega_0^2}{4} e^{-3Ht},$$

which combine to yield

$$\frac{d\bar{\rho}_\phi}{dt} \approx -3H\bar{\rho}_\phi. \tag{13.45}$$

In contrast to the nearly constant energy density that was present throughout the slow-roll phase of inflation, the energy density evolves during this oscillating stage in the same way that nonrelativistic matter does. To put this another way, as ϕ oscillates around the minimum of the potential, the universe is effectively filled with a condensate of massive, pressureless particles.

As a final step, we can introduce couplings between the inflaton and the particles of the Standard Model (or to other particles that themselves couple to the Standard Model). Such couplings can enable the inflaton to decay, causing the energy density to evolve as follows:

$$\frac{d\bar{\rho}_\phi}{dt} \approx -3H\bar{\rho}_\phi - \Gamma_\phi \bar{\rho}_\phi, \tag{13.46}$$
$$\frac{d\rho_{\rm SM}}{dt} \approx -4H\rho_{\rm SM} + \Gamma_\phi \bar{\rho}_\phi,$$

where Γ_ϕ is the decay width of the inflaton.

As the inflatons decay, their energy density is transferred into a bath of relativistic Standard Model particles. Although the resulting temperature of this bath is quite model-dependent, we can draw some general conclusions about this quantity. First of all, the energy density of the bath after reheating must be less than the energy density present during inflation. Furthermore, in order for the Standard Model particles to reach a state of thermal equilibrium, it must be the case that $\Gamma_\phi \gtrsim H$. From this, one finds that the reheating temperature must satisfy the following condition [211]:

$$\Gamma_\phi \gtrsim \left(\frac{8\pi\rho_{\rm SM}}{3M_{\rm Pl}^2}\right)^{1/2} = \left(\frac{8\pi^3 g_\star(T_{\rm RH})T_{\rm RH}^4}{90M_{\rm Pl}^2}\right)^{1/2}, \tag{13.47}$$

and thus

$$T_{\rm RH} \lesssim 7 \times 10^{13}\ {\rm GeV} \times \left(\frac{106.75}{g_\star(T_{\rm RH})}\right)^{1/4} \left(\frac{\Gamma_\phi}{10^{10}\ {\rm GeV}}\right)^{1/2}. \tag{13.48}$$

It is, of course, possible that the reheating temperature could have been much lower than this upper limit. The only strict lower limit on this temperature is set by the requirement that BBN take place normally, in order to accommodate the measured light element abundances. From this, we conclude that the reheating temperature of our universe must have been at least a few MeV.

13.8 Present and future probes of inflation

When inflation was originally proposed, the only empirical evidence for such an era was the approximate flatness and uniformity of our universe, along with the lack of any detectable heavy relics. Over the past four decades, however, inflation has scored some important empirical successes. First of all, it has become clear over this time that our universe is not merely approximately flat, but is extremely flat, $\Omega_\kappa \equiv 1 - \Omega_{\rm tot} = 0.0007 \pm 0.0019$. Furthermore, measurements of the CMB anisotropies have revealed the spectrum of primordial density perturbations to be approximately scale-invariant, consistent with expectations from inflation. More quantitatively, the measured values of $n_s = 0.9649 \pm 0.0042$ and $dn_s/d\ln k = -0.0045 \pm 0.0067$ are each compatible with a wide range of inflationary models.

The lack of tensor modes detected in the CMB, currently measured at a level of $r \lesssim 0.032$, allows us to rule out a wide range of large-scale inflation models, requiring $V \lesssim (10^{16}\ {\rm GeV})^4$ when evaluated at $\phi_{\rm CMB}$. Roughly speaking, this favors inflationary potentials that are approximately flat on the scales over which the observed primordial density perturbations are generated, and that later steepen to bring inflation to an end.

To our current ability to measure, the CMB's temperature distribution appears to be Gaussian. While this is expected within the context of inflation, many models predict that small non-Gaussian signatures will be encoded within the CMB, as well as within the large-scale structure of our universe. Future measurements that constrain or characterize these non-Gaussianities will be important in discriminating between various models of inflation.

13.9 Eternal inflation

Before bringing this chapter to an end, I want to point out one particularly surprising feature of many inflationary models: namely, that inflation is predicted to never end [212, 213] (see also Ref. [214]).

To understand what I mean by this, consider a region of space that is undergoing inflation, with a radius equal to the Hubble length. Over one Hubble time, the scale factor will increase by a factor of e, and thus the volume of the region in question will increase by a factor of $e^3 \sim 20$. Since the energy density, and thus the Hubble length, is not significantly changing during inflation, this region has grown to become approximately 20 different regions, each of which is causally disconnected from the others. Furthermore, over the course of this Hubble time, the scalar field advances down the slope of the potential according to the classical equation of motion (Eq. 13.13), changing in value by $\Delta\phi_{\rm cl} \approx \dot{\phi}_{\rm cl}/H$.

This evolution, however, is not a purely classical phenomenon, and thus is also impacted by quantum fluctuations. This leads to a change in the value of the scalar field over a Hubble time that is given by

$$\begin{aligned}\Delta\phi &= \Delta\phi_{\rm cl} + \Delta\phi_{\rm qu} \\ &\approx \frac{\dot{\phi}_{\rm cl}}{H} + \Delta\phi_{\rm qu},\end{aligned} \tag{13.49}$$

where the quantum fluctuations, $\Delta\phi_{\rm qu}$, are drawn from a Gaussian distribution centered at zero and with a width of $\sim H/2\pi$ [204, 215, 216, 217].

In each causally disconnected region, the value of $\Delta\phi_{\rm qu}$ is randomly and independently selected. Since $\Delta\phi_{\rm cl}$ is negative and the average value of $\Delta\phi_{\rm qu}$ is zero, the scalar field in the median region will continue to roll down the slope of the potential, allowing inflation to proceed normally and ultimately come to an end. In a fraction of these regions, however, the quantum fluctuation, $\Delta\phi_{\rm qu}$, will be positive and greater in magnitude than $\Delta\phi_{\rm cl}$, causing the total value of the scalar field to *increase* over the course of the Hubble time. If this occurs with a sufficiently low probability, the quantum fluctuations will be of little consequence. But if the probability of $\Delta\phi$ being positive in a given region is larger than $e^{-3} \sim 0.05$, then the total number of regions with a particular value of ϕ will tend to *increase* as time advances. In other words, if this condition is met, then the volume of space that is at a given stage of inflation will grow with time.

To see whether this condition is met in a given model of inflation, we note that in drawing from a Gaussian distribution centered around zero, there is a 5% chance that the value will be positive and more than 1.64 standard deviations from the mean. This implies that the volume of inflating space will tend to increase if

$$\frac{\dot{\phi}_{\rm cl}}{H} + 1.64\,\frac{H}{2\pi} > 0, \tag{13.50}$$

corresponding to $H^2/|\dot{\phi}_{\rm cl}| > 3.8$. Using the classical equation of motion to relate $\dot{\phi}_{\rm cl} \approx -(dV/d\phi)/3H$ and the second Friedmann equation to write $H^2 = 8\pi V/3M_{\rm Pl}^2$, this condition can be expressed as follows:

$$\frac{V^{3/2}}{|dV/d\phi|} \gtrsim 0.052\, M_{\rm Pl}^3. \tag{13.51}$$

As it turns out, this condition is generically satisfied across a wide range of inflation ary models.

To be clear, just because inflation continues across an ever increasing volume of space, this does not mean that it does not end somewhere. In fact, in the most regions of space, the scalar field moves down the inflationary potential in each Hubble time, evolving approximately as predicted by the classical equation of motion. Thus, throughout most of the inflating space, inflation proceeds normally. As the volume of inflating space grows exponentially, however, the total volume of inflating space can increase monotonically, leading to the unbounded growth of the number of causally disconnected regions in which inflation is occurring. When inflation ends at a given point in space, the result is a spatially distinct and causally disconnected region filled with matter and energy—what I would call a "universe." Thus, if the condition described above is satisfied, inflation will lead to the creation of an infinite number of universes, or what is known as the inflationary multiverse.

Additional reading

For a remarkably clear review of inflation, I recommend Baumann's TASI lectures [218] (see also Ref. [219]). I have also found the Particle Data Group's review of this subject [24] to be helpful.

Problems

13.1 Assuming that the early universe was consistently radiation-dominated, what is the largest degree of curvature that could have been present during the era of grand unification ($T \sim 10^{16}$ GeV) or the Planck era ($T \sim 10^{19}$ GeV), without conflicting with the measurement of $|\Omega_\kappa| \lesssim 0.02$ today?

13.2 Consider our current universe with $\Omega_\Lambda \approx 0.685$ and $\Omega_M \approx 0.315$. Is the magnitude of large-scale curvature increasing or decreasing at this time? At what value of the scale factor did (or will) our universe transition from becoming more curved to becoming less so?

13.3 Consider a stable state with a mass near the scale of grand unification, $M \sim 10^{16}$ GeV. If these particles fell out of equilibrium at $T \sim 10^{16}$ GeV, what would their abundance be today (if not for inflation)?

13.4 If instead of standard inflation, imagine that there had been an era in the early universe during which the equation of state was $w = -0.5$. Roughly how long (in terms of scale factor) must this state have persisted in order to solve the flatness and horizon problems?

13.5 Express the Friedmann equations using reduced Planck units.

13.6 Using Eqs. 13.11 and 13.13, show that $\dot{H} = -8\pi G\dot{\phi}^2/2$.

13.7 Starting with Eq. 13.14, derive Eq. 13.18.

13.8 Show that the two expressions for the slow-roll parameter, η, given in Eq. 13.19 are equivalent.

13.9 Under the slow-roll conditions of $|\ddot{\phi}| \ll |3H\dot{\phi}|, |dV/d\phi|$ and $V \gg \dot{\phi}^2/2$, show that $\epsilon_v \approx \epsilon$.

13.10 Under the condition of $V \gg \dot{\phi}^2/2$, show that $\eta \approx \eta_v - \epsilon_v$.

13.11 For the case of $V = m_\phi^2\phi^2/2$, calculate the value of the inflaton mass, m_ϕ, that yields the measured value of A_s.

13.12 For an inflationary potential of the form $V \propto \phi^4$, calculate n_s, $dn_s/d\ln k$, and r.

13.13 Generalize the previous problem for the case of an inflationary potential of the form $V \propto \phi^n$, calculating n_s, $dn_s/d\ln k$, and r.

13.14 Consider an inflationary potential of the following "hilltop" form [220]:

$$V(\phi) = V_0\left[1 - \left(\frac{\phi}{\phi_0}\right)^4\right].$$

Show that for this potential, it follows that

$$r = \frac{8}{3}\frac{(1-n_s)}{\left[1 + \frac{(1-x^4)}{2x^4}\right]},$$

where $x \equiv \phi_{\rm CMB}/\phi_0$.

13.15 In the model known as "natural inflation" [221, 222], the inflaton potential is given by

$$V \propto 1 + \cos\left(\frac{\phi}{\phi_0}\right).$$

For the choice of $\phi_0 = 2M_{\rm Pl}$, calculate n_s and r. Note that this problem is most easily done numerically.

13.16 Consider the inflationary potential $V = V_0\,\phi^4$. Is this model of inflation eternal? More specifically, determine the minimum value of V for which the requirements of eternal inflation are satisfied.

14

Dark Energy

One of the most surprising discoveries in all of modern cosmology has been that of dark energy. The fact that our universe is expanding at an accelerating rate implies that it has recently entered an era in which the energy density is dominated by something that does not appreciably dilute as space expands. Whether this dark energy is a cosmological constant or is instead the consequence of some kind of evolving scalar field remains an open question. In either case, there remains much that we do not yet understand about this mysterious substance.

14.1 The cosmological constant problem

As originally conceived by Einstein, general relativity is a purely classical theory. In this context, the cosmological constant, Λ, is simply a free parameter, and there is nothing about the structure of general relativity itself that gives us any insight into the value of this quantity. If we take quantum mechanics into account, however, this question takes on a somewhat different character. In particular, we should expect the energy density of the vacuum to receive contributions from the quantum fluctuations of each of the fields that exists in nature. After all, each mode of a free quantum field acts like a simple harmonic oscillator, whose ground-state energy is nonzero and positive.

When we attempt to calculate the energy density of the quantum vacuum, we run into serious problems. In particular, when we perform an integral over all of the modes of a given field, we find that the result diverges, leading to an infinite energy density. Our only hope of avoiding this nonsensical conclusion is to propose that, for some yet unknown reason, the shortest length scale (i.e., highest energy) modes do not contribute to the energy density, allowing us to truncate the integral and obtain a finite result.

Even with such a cutoff, however, we still encounter what seems to be a fatal problem. Namely, this calculation yields an energy density that is on the order of the cutoff energy to the fourth power. Despite being finite, this result is far too large to be compatible with the observed features of our universe. For example, if we were to speculate that, in the correct theory of quantum gravity, only modes longer than the Planck length contribute to the energy density of the vacuum, we would obtain an estimate of $\rho_\Lambda \sim M_{\rm Pl}^4 \sim 2 \times 10^{76}\,{\rm GeV}^4 \sim 6 \times 10^{122}\,\rho_{\rm crit}$. This is obviously far too large. To reduce the magnitude of this prediction, one might propose that some kind of new physics come into play at a lower energy scale, allowing us to truncate this integral more severely (see, for example, Ref. [223, 224]). If we

posit that the particles and fields associated with this new physics lie just beyond the reach of the LHC, for example, we would estimate $\rho_\Lambda \sim (3\,\text{TeV})^4 \sim 2 \times 10^{60}\,\rho_{\text{crit}}$. This too does not provide a satisfactory resolution to the problem at hand.

Throughout most of the twentieth century, it was widely speculated that some unknown symmetry or other mechanism must somehow be built into the laws of physics that forces the energy density of the vacuum to exactly vanish, $\rho_\Lambda = 0$. While no such physics has ever been identified, this possibility seemed to many to be at least plausible. This too was turned on its head, when it was discovered in the 1990s that our universe's expansion rate is accelerating. This surprising fact seems to imply that the energy density of the vacuum is very small, and yet finite, raising more questions than it offers answers, and leaving us in a generally confused state regarding the quantum nature of our universe's vacuum.

14.2 Dark energy and accelerating expansion

In any universe that is dominated by either matter or radiation, the expansion rate decreases with time, $\ddot{a} < 0$. The discovery that our universe's expansion rate is accelerating thus implies that its energy density must be dominated by an exotic component, with a negative equation of state, $w < 0$ (see Secs. 2.5 and 2.6). This exotic energy could take the form of a positive cosmological constant, or of some other kind of energy that is approximately uniform in both space and time. To capture both of these possibilities, this substance is often referred to simply as "dark energy."

The first direct evidence that our universe's expansion rate is accelerating came from measurements of the evolving Hubble rate using type Ia supernovae [225, 226]. Within only a few years, this conclusion was further supported by measurements of the first peak in the angular power spectrum of the CMB, whose location indicates that $\Omega_{\text{tot}} \approx 1$ (see Sec. 7.6.1). Given that Ω_M was already known to be much less than unity, this too appeared to require the presence of a positive cosmological constant.

Modern measurements of the Hubble rate, CMB, and large-scale structure indicate that $\rho_\Lambda \approx 5.86 \times 10^{-33}\,\text{kg/cm}^3$, corresponding to $\Omega_\Lambda \approx 0.685$. While this is large enough to dominate the energy density of our universe, it is $\sim 10^{123}$ times smaller than the estimate made in the previous section (for a cutoff near the Planck scale). Nothing we know about the laws of physics seems to offer us any insight into the value of ρ_Λ.

The discovery of accelerating expansion has also raised a question known as the coincidence problem. Whereas the cosmological constant problem asks why the energy density of the vacuum is not much larger than it is observed to be, modern cosmological observations also force us to contend with the question of why ρ_Λ and ρ_M are of the same order of magnitude in the present era. Alternatively, this can be phrased to ask why it is that we find ourselves living in the very special time when the vacuum energy has only recently begun to dominate the energy density of our universe. With the possible exception of anthropic selection (see Sec. 14.5), I am not aware of any particularly satisfactory answers to this question.

14.3 Dynamical dark energy

Instead of being described by a cosmological constant, it is possible that the vacuum energy of our universe evolves with time. Motivated by the challenges that we face in understanding the energy density of our universe's vacuum, many different models of dynamical dark

energy have been proposed. Such models of dynamical dark energy, or quintessence, often rely on a slowly rolling scalar field, in close analogy to inflation [227, 228, 229, 230, 231, 232].

Much like inflation, the equation of state of a scalar field responsible for dark energy will be approximately $w \approx -1$ so long as its potential energy is much greater than its kinetic energy, $V(\phi) \gg \dot{\phi}^2/2$ (see Eq. 13.12). In order for the energy density associated with this scalar field to have remained approximately constant throughout the last few billion years (enabling it to effectively mimic a cosmological constant), the value of this field must not be evolving too rapidly. In other words, the slow-roll conditions that we are familiar with from our discussion of inflation must be satisfied in this case as well. After relating the mass of the scalar field to the second derivative of the potential, $m_\phi^2 = d^2V/d\phi^2$, the slow-roll condition of $\eta_v < 1$ can be rewritten as $m_\phi^2 < 8\pi V/M_{\rm Pl}^2$. Using the second Friedmann equation, this is equivalent to $m_\phi \lesssim \sqrt{3}\, H_0$. Thus any scalar field that could potentially be responsible for dark energy must have an exceedingly small mass, $m_\phi \lesssim 10^{-33}$ eV. Any heavier scalar fields would have rolled into the minimum of their potentials long ago.

There are a number of challenges that one must contend with when proposing the existence of a very light scalar field. In particular, we expect the mass of a scalar field to receive radiative corrections, which will tend to drive its value upward (see Sec. 17.1.1). From this perspective, it is hard to understand how any scalar field could be light enough to provide a viable explanation for the dark energy of our universe. Furthermore, a very light scalar field would be expected to induce long-range forces and lead to time-varying coupling constants, none of which have been observed in our universe [233, 234]. We will discuss these challenges further in Sec. 14.3.2.

14.3.1 An example of quintessence

As an example of a model of quintessence, consider a scalar field with a potential given by $V = V_0\, \phi^{-2}$ [228, 231]. During the radiation-dominated era, the energy density associated with this scalar field was small, allowing us to describe the time evolution of the Hubble rate in the standard way, $H = 1/2t$. Inserting this into the scalar field's equation of motion (see Eq. 13.13), we have

$$\ddot{\phi} + 3H\dot{\phi} + \frac{dV}{d\phi} = 0, \tag{14.1}$$

$$\ddot{\phi} + \frac{3}{2t}\dot{\phi} - \frac{2V_0}{\phi^3} = 0.$$

One solution to this equation of motion is given by

$$\phi = \sqrt{2}\, V_0^{1/4}\, t^{1/2}. \tag{14.2}$$

During matter domination, with $H = 2/3t$, the time dependence of this solution remains unchanged, while the normalization shifts slightly, to $\phi = (8V_0/3)^{1/4}\, t^{1/2}$. While this is only one of the possible solutions to the equation of motion, it is an example of what is known as an "attractor solution." By this I mean that the other solutions tend to evolve with time toward this attractor solution. As a result, the behavior of this model is relatively insensitive to the initial conditions of the problem.

At later times, when the scalar field dominates the energy density of the universe, the equation of motion is given by

$$\ddot{\phi} + 3\left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2}\left(\frac{\dot{\phi}^2}{2} + \frac{V_0}{\phi^2}\right)^{1/2}\dot{\phi} - \frac{2V_0}{\phi^3} = 0. \tag{14.3}$$

If the potential is large enough to efficiently damp the evolution of ϕ, this becomes

$$3\left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2}\left(\frac{V_0}{\phi^2}\right)^{1/2}\dot{\phi} - \frac{2V_0}{\phi^3} \approx 0, \tag{14.4}$$

whose solution is $\phi = (3V_0 M_{\rm Pl}^2/2\pi)^{1/6}\, t^{1/3}$.

We are now in a position to calculate the Hubble rate at late times in this model. Using the expression for ϕ given above, we can write $\rho \approx V \approx V_0\phi^{-2} \approx (2\pi\, V_0^2/3M_{\rm Pl}^2)^{1/3}\, t^{-2/3}$, and $H = (32\pi^2 V_0/9M_{\rm Pl}^4)^{1/3}\, t^{-1/3}$. Solving the first Friedmann equation for this Hubble rate, we find that the scale factor in this case evolves as

$$a \propto e^{(t/\tau)^{2/3}}, \tag{14.5}$$

where $\tau = M_{\rm Pl}^2/(2\sqrt{3}\,\pi\, V_0^{1/2})$. Note that this result is similar to, but not quite as fast as, that predicted for a universe that is dominated by a cosmological constant, $a \propto e^{t/\tau}$.

14.3.2 Other models of dynamical dark energy

Quintessence does not provide us with an explanation for the coincidence problem of dark energy. Furthermore, as mentioned earlier, we expect the scalar field in quintessence models to couple to Standard Model particles, at least to some degree. Even if those couplings are suppressed by the Planck scale, the resulting interactions should generically lead to the appearance of observable long-range forces and time-dependent coupling constants. These challenges have motivated a great deal of investigation into models of dynamical dark energy that could potentially address these and other outstanding problems.

One example is a class of models known as *k*-essence [235]. In these models, the standard (or "canonical") expression for the kinetic energy of a scalar field, $\dot{\phi}^2/2$, is replaced by another function of ϕ and $\dot{\phi}$. For some choices of this function, the value of the scalar field can evolve at a rate that depends on the equation of state of the background. More specifically, it can be possible in some cases to obtain solutions in which $\rho_\phi \propto \rho_{\rm tot}$ during radiation domination, that then transition to a state in which ρ_ϕ is approximately constant once matter domination begins. The appeal of these models derives from their ability to potentially offer an explanation for why typical observers would find themselves living in a universe that contains comparable quantities of matter and dark energy. Such models, however, also appear to require a great deal of fine-tuning [236], significantly dampening enthusiasm for such ideas.

Another class of such models involves a scalar field with a periodic potential. As an example, consider a scalar field with the following potential:

$$V = V_0\, e^{-\phi/\phi_0}\left[1 + \alpha\cos\left(\frac{\phi}{\phi_1}\right)\right], \tag{14.6}$$

where V_0, ϕ_0, α, and ϕ_1 are each free parameters. The exponential part of this potential, if taken alone, would lead the field to evolve as $\phi \propto \ln t$, and to $\rho_\phi \propto \rho_{\rm tot}$. When the sinusoidal

factor is included, the density of the scalar field still tracks the overall energy density of the universe when averaged over large periods of time, but exhibits oscillatory behavior on shorter timescales [237]. This causes the universe to go through alternating periods of matter and dark energy domination, with a frequency that is set by the rate of Hubble expansion. Thus, within the context of this model, it is perhaps not particularly surprising that we find ourselves living near the beginning of an era that is dominated by dark energy.

Other models of dynamical dark energy have been motivated by the problem of the long-range forces that are expected to be induced by light scalar fields. In order to satisfy the slow-roll conditions, any scalar field responsible for dark energy must be exceedingly light, $m_\phi \lesssim 10^{-33}$ eV. Such particles should communicate forces across distances as great as $d \sim m^{-1} \gtrsim 10^{10}$ pc, and with a strength that is no weaker than that of gravity. As a consequence, both solar system and laboratory constraints on "fifth forces" provide a major challenge for models of dynamical dark energy.

One possible way to evade this class of constraints is to consider a scalar field that couples to baryons in such a way as to cause its mass to increase in the presence of ordinary matter. So-called "chameleon" fields [238, 239] could be very light throughout the vast majority of our universe's volume, thus satisfying the slow-roll conditions required of dark energy, while being much more massive in the vicinity of the solar system.

14.4 Observational probes of dark energy

A variety of observational probes have been pursued in an effort to measure how the density of dark energy has evolved over cosmic time, enabling us to constrain the degree to which dark energy might be dynamical rather than a cosmological constant. These measurements are each sensitive to the equation of state of dark energy as a function of redshift, $w(z)$. This function, in turn, can be expressed in terms of either the angular diameter distance, or the luminosity distance, as a function of redshift. In particular, for the case of a flat universe, we can use Eqs. 2.57 and 2.59, along with the first Friedmann equation, to write the angular diameter distance as follows:

$$\begin{aligned} d_A(z) &= \frac{1}{1+z} \int_a^1 \frac{da'}{a'^2\, H(a')} \\ &= \frac{1}{1+z} \left(\frac{3M_{\rm Pl}^2}{8\pi}\right)^{1/2} \int_{1/(1+z)}^1 \frac{da'}{a'^2\, [\rho_{DE}(a') + \Omega_{M,0} a'^{-3} + \Omega_{\rm rad,0}\, a'^{-4}]^{1/2}}. \end{aligned} \tag{14.7}$$

Note that the density of dark energy in this expression is explicitly a function of redshift (or, equivalently, scale factor). By measuring the function, $d_A(z)$, we can infer the evolution of the density (and thus the equation of state) of dark energy.

Observations of the CMB's angular power spectrum provide us with a measurement of our universe's total energy density, as well as of the densities of dark matter and baryonic matter. From this information, it is clear that most of the energy in our universe does not consist of matter, requiring the presence of dark energy to make up this difference, $\Omega_\Lambda = 0.6847 \pm 0.0073$. The CMB, however, does not tell us very much about the detailed nature of dark energy, or how its density has evolved over the course of cosmic history.[1]

[1] Exceptions to this are the secondary anisotropies in the CMB, such as those induced through the integrated Sachs-Wolfe effect, the Sunyaev-Zel'dovich effect, and gravitational lensing (see Secs. 7.5.2 and 7.5.5). As these secondary anisotropies are sensitive to the low-redshift large-scale structure of our universe, they provide us with a valuable probe of how the density of dark energy has evolved over time.

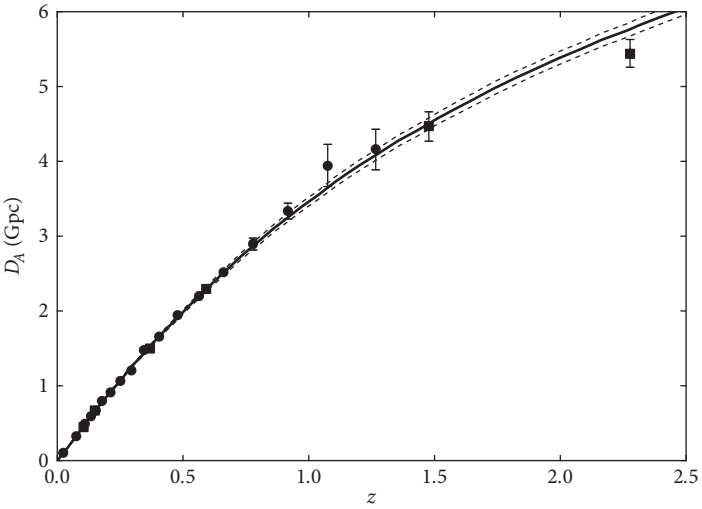


Figure 14.1 Measurements of the comoving angular diameter distance as a function of redshift from Type Ia supernovae (circles) [241] and baryon acoustic oscillations (squares) [245, 246, 247, 248]. The solid line is the prediction for a cosmological constant ($w = -1$), while the dashed lines correspond to dark energy with an equation of state of $w = -1.1$ (upper) or $w = -0.9$ (lower).

Some of the most powerful probes of dark energy have been based on direct measurements of the Hubble rate, $H(z)$. Through observations of Type Ia supernovae [240, 241, 242, 243, 244], the Hubble rate has been measured out to redshifts as large as $z \approx 1.4$. Type Ia supernovae are particularly useful in this context because they are very bright and exhibit peak luminosities that can be predicted to an accuracy of 10–15% after accounting for the duration and color of their light curves. This allows us to use these objects as standard (or standardizable) candles.

As discussed in Sec. 10.7, measurements of the baryon acoustic oscillations in the matter power spectrum can be used to determine the length scale of the sound horizon. This information, in turn, allows us to constrain how our universe has expanded since the time of recombination. Through these measurements, baryon acoustic oscillations provide us with a valuable probe of cosmological parameters, especially those associated with dark energy.

Recall from Sec. 10.3 that dark energy's influence on the expansion rate suppresses the growth of large-scale structure. For this reason, measurements of our universe's matter distribution provide us with information that we can use to constrain the nature of dark energy. Among other probes of our universe's evolving large-scale structure, this information comes from measurements of the distribution of galaxy clusters, and from the weak gravitational lensing of distant galaxies by the more nearby and clustered distribution of matter.

In Figure 14.1, measurements are shown of the comoving angular diameter distance, $d_A \times (1 + z)$, as a function of redshift, based on observations of Type Ia supernovae [241] and baryon acoustic oscillations [245, 246, 247, 248]. The solid line is the prediction for a model with dark energy in the form of a cosmological constant ($w = -1$), for the best-fit values of Ω_M and H_0 and assuming $\Omega_{\rm tot} = 1$. The dashed lines demonstrate how this prediction changes if the dark energy's equation of state is increased or decreased by 10%. The existing data are consistent with the dark energy behaving as a cosmological constant. More specifically, if the dark energy's equation of state is assumed to be constant, current observations allow us to constrain this quantity to lie within a few percent of -1 [247, 8].

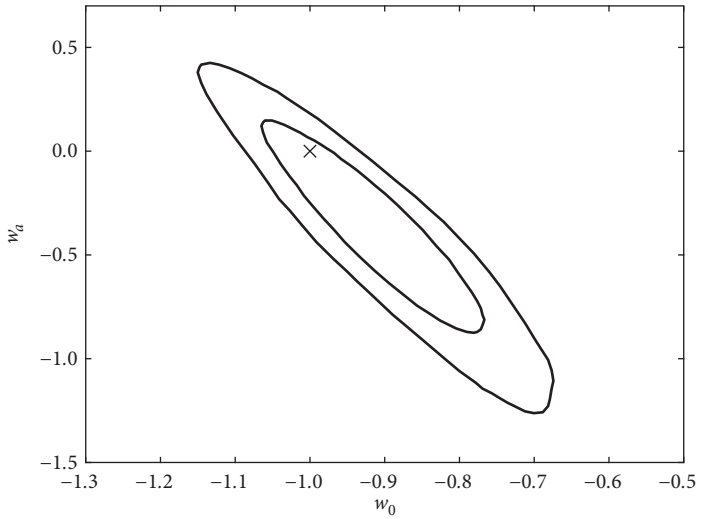


Figure 14.2 Constraints on the dark energy's evolving equation of state, for a parameterization given by $w(a) = w_0 + w_a\,(1 - a)$. The $\times$ symbol represents the case of a cosmological constant.

If the dark energy's equation of state in not a constant but instead evolves with redshift, these constraints can be significantly relaxed. Constraints on such models are often presented in terms of the coefficients in the following linear expansion of the dark energy's evolving equation of state:

$$w(a) = w_0 + w_a\,(1 - a), \tag{14.8}$$

where w_0 is dark energy's equation of state at the present time, and w_a characterizes its recent evolution. The current constraints on these parameters are shown in Figure 14.2, assuming no spatial curvature. These constraints are based on a combination of measurements including those of the CMB, baryon acoustic oscillations, the distribution of galaxies, and Type Ia supernovae [247]. While these results are entirely consistent with dark energy in the form of a cosmological constant ($w_0 = -1$, $w_a = 0$), a nonnegligible degree of evolution remains a possibility.

14.5 Dark energy and the anthropic principle

It is possible that some yet to be understood laws of physics are responsible for dark energy. Perhaps if we were to discover these laws, they would enable us to determine why the density of this mysterious substance is equal to the value we have measured. It is also possible, however, that the abundance of dark energy is something that cannot be calculated from first principles. Instead, the density of dark energy could vary across very large distance scales (larger than the size of the cosmological horizon), or even from universe to universe. If this is the case, it would be pointless for us to try to deduce the energy density of the local vacuum. Instead, we would be better served by asking questions along the lines of, "what would a typical observer measure the energy density of the vacuum to be?"

If this kind of reasoning is new to you, it might seem strange that we would need to make any reference at all to a "typical observer." After all, why would the physical characteristics of a universe (or multiverse) have anything to do with the distribution of observers it contains? It is entirely plausible, however, that life—and thus observers—can only exist within a small

fraction of the multiverse. If this is the case, then we will necessarily find ourselves in an exceptional location, with rare physical conditions. This reasoning is an application of what is known as the "anthropic principle" [249], and it has come to play a major role in how many cosmologists think about dark energy.

Anthropic reasoning as applied to the cosmological constant problem goes back to Steven Weinberg [250], who made the following argument years *before* the discovery of dark energy. The energy density of the vacuum depends on the detailed properties of all of the fields that exist (as described in Sec. 14.1). If these fields vary from universe to universe, then we should expect the energy density of the vacuum to be on the order of $\rho_\Lambda \sim M_{\rm Pl}^4 \sim 2 \times 10^{76}\ {\rm GeV}^4 \sim 6 \times 10^{122}\ \rho_{\rm crit}$ in most universes, but not necessarily in all. In a small fraction of universes, cancellations will occur between different terms, leading to a lower—and perhaps much lower—value of ρ_Λ. With this in mind, let's ask ourselves what value of ρ_Λ we should expect a typical observer to measure. In an average universe, with $\rho_\Lambda \sim 6 \times 10^{122}\ \rho_{\rm crit}$, vacuum energy becomes dominant around the Planck time, and exponential expansion quickly dilutes away any matter or radiation that might exist. The more we lower the value of ρ_Λ, the longer we delay the transition to vacuum energy domination and exponential expansion. If we reduce it to $\rho_\Lambda \sim 10^3\ \rho_{\rm crit}$, for example, vacuum energy domination would begin a few hundred million years after the Big Bang, which is roughly around the time that the first galaxies and stars were beginning to form. We can thus see why all observers would find themselves living in a universe in which ρ_Λ is not much larger than $\sim 10^3\ \rho_{\rm crit}$, despite the fact that universes with such a small value of ρ_Λ would be extremely rare throughout the multiverse. Taking this reasoning further, the smaller that ρ_Λ is, the longer galaxy formation will be able to continue, leading to the production of more stars and increasing the probability that life will emerge. If we assume that the probability of life appearing in a given universe is roughly proportional to the total number of stars that are produced in that universe, then it follows from this argument that a typical observer should expect to find themselves living in a universe with a vacuum energy density that is not very different from that measured in our universe [251, 252].

It should be noted that, whereas quintessence and other dynamical explanations for dark energy do not generally attempt to explain why there is not a large cosmological constant in the first place (on top of any energy density associated with the evolving scalar field), the anthropic approach does provide a potential explanation for the cosmological constant problem. Namely, the average value of the cosmological constant is much larger than the value measured by the average observer.

Additional reading

For further discussion of the cosmological constant problem, I recommend the TASI lectures on this subject by Bousso [253] and Carroll [254]. For a more detailed treatment of dynamical dark energy models, I like the review by Copeland, Sami, and Tsujikawa [255]. For more information regarding the observational status of dark energy, I suggest the Particle Data Group's review of this subject [24].

Problems

14.1 What if instead of estimating the energy density of the vacuum to be $\rho_\Lambda \sim M_{\rm Pl}^4$, I had used the expression $\rho_\Lambda = \Lambda M_{\rm Pl}^2/8\pi$ and estimated Λ to be the inverse square of the Planck length. How would this impact our conclusions?

14.2 In the opening section of this chapter, we estimated that $\rho_\Lambda \sim M_{\rm Pl}^4 \sim 2 \times 10^{76}\ {\rm GeV}^4 \sim 6 \times 10^{122}\ \rho_{\rm crit}$. Can you think of any calculation that you (or anyone else) has ever done that has been wrong by a factor as large as this?

14.3 In a universe with $\rho_\Lambda \sim M_{\rm Pl}^4$, calculate how long after the Big Bang the cosmological constant begins to dominate the total energy density. What about if $\rho_\Lambda \sim (3\ {\rm TeV})^4$?

14.4 Instead of $V = V_0\phi^{-2}$, consider a model of quintessence with a general power-law potential of the form $V = V_0\phi^{-\alpha}$. Find the solution to the equation of motion during the radiation-dominated era, in analogy to Eq. 14.2.

14.5 Show that Eq. 14.2 is, in fact, a tracker solution by determining how a small perturbation around that solution will evolve with time.

14.6 Show that, in the large t limit, the approximation $\dot{\phi}^2/2 \ll V$ used in deriving Eq. 14.4 is valid for the solution $\phi = (3V_0 M_{\rm Pl}^2/2\pi)^{1/6}\, t^{1/3}$.

14.7 Consider once again a model of quintessence with a general power-law potential $V = V_0\phi^{-\alpha}$. Find how the scale factor evolves with time in the late-time and scalar field-dominated limit.

14.8 For a scalar field with a potential given by $V = V_0\, e^{-\phi/\phi_0}$, determine how ϕ evolves with time during the radiation-, matter-, and scalar field-dominated eras. Show that the energy density of this scalar field is proportional to the total energy density of the universe throughout the radiation- and matter-dominated eras.

14.9 Consider a universe that is identical to ours, except that it has a different value of ρ_Λ. As a function of this parameter, calculate how long after the Big Bang it is before $\rho_\Lambda = \rho_M$.

14.10 Consider an ensemble of universes that are like ours, except that the value of ρ_Λ is distributed uniformly between 0 and $6 \times 10^{122}\ \rho_{\rm crit}$. Further assume (somewhat unrealistically) that the star formation rate in a given universe is zero for $t < t_0 = 10^9$ years, positive and constant after $t = t_0$, and zero again once $\rho_\Lambda > \rho_M$. Calculate the value of ρ_Λ for the universe in which the median star forms.

III

Particle Astrophysics

15

Cosmic Rays

Astrophysical objects accelerate charged particles to incredible energies. These so-called "cosmic rays" are a subject of intense study, and reveal much to us about our universe's most extreme astrophysical environments. In this chapter, I summarize what we know about the spectrum and composition of the cosmic rays and describe how these particles propagate and interact as they travel through our galaxy and universe. I also discuss the process of cosmic-ray acceleration, and outline what we have been able to deduce regarding the origin of the cosmic-ray spectrum. The material presented in this chapter will be important in our later discussions of gamma-ray and neutrino astronomy, as well as indirect searches for dark matter, as covered in Chapters 16 and 19.

15.1 The cosmic-ray spectrum

The cosmic-ray spectrum extends from a few MeV up to energies greater than 10^{11} GeV. At low and intermediate energies, the composition of the cosmic-ray spectrum is dominated by protons, along with smaller but significant numbers of electrons and various species of nuclei. In fact, all nuclei with a lifetime of at least 10^6 years are found within the cosmic-ray spectrum. In addition, positrons and antiprotons are also observed among the cosmic rays. Due to the effects of the galactic magnetic field, the arrival directions of these particles are nearly isotropic.

The shape and composition of the cosmic-ray spectrum provide us with a number of clues about where these particles may have been accelerated, and how they propagate through the Milky Way. At energies between several GeV and 100 TeV, the cosmic-ray spectrum is well described by a power-law with a spectral index of -2.7, $dN/dE \propto E^{-2.7}$. At these energies, it is generally agreed that the cosmic rays originate mostly from galactic sources, including supernova remnants.

At relatively low energies, cosmic rays can be detected directly using instruments on board high-altitude balloons or satellites. At higher energies, these particles are instead studied by measuring the air showers that are generated when they collide with the top of earth's atmosphere. Such interactions produce a large number of pions, which decay to generate photons, electrons, and positrons, as well as muons, all of which can be detected on

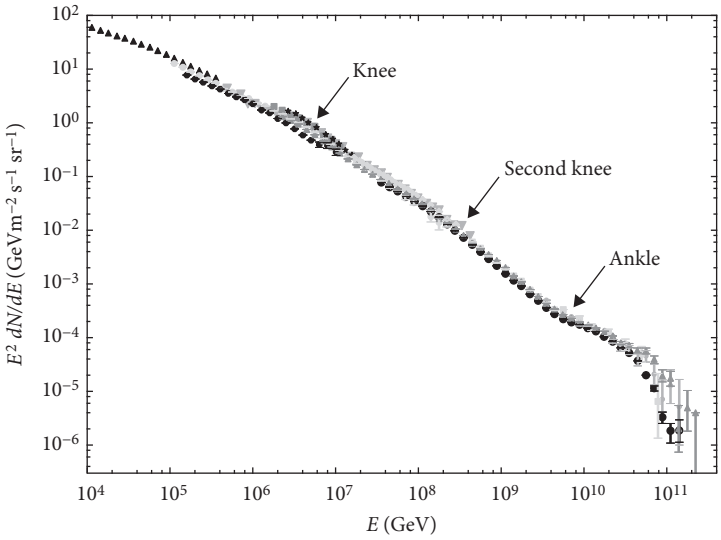


Figure 15.1 The all-particle cosmic-ray spectrum, as measured by various air shower experiments [256, 257, 258, 259, 260, 261, 262, 263, 264, 265].

earth's surface. In addition, particularly high-energy air showers ($E \gtrsim 10^8$ GeV) can produce detectable amounts of fluorescence that can be measured using complementary telescopes.

The cosmic-ray spectrum is shown in Figure 15.1, as measured by various air shower experiments including HAWC [256], MSU [257], Tibet [258], Akeno [259], CASA-MIA [260], Kascade [261], Kascade Grande [262], HiRes [263], Auger [264], and Telescope Array [265]. This spectrum includes features known as the "knee," the "second knee," and the "ankle." While interpretations of these features are still debated, the steepening of the spectrum associated with the knee appears to be the consequence of galactic cosmic-ray sources being unable to accelerate protons to energies above $\sim 10^6$ GeV [266]. In this energy range, heavier nuclei accelerated by galactic sources are instead expected to dominate the cosmic-ray spectrum (we'll return to the maximum energy of cosmic-ray acceleration in Sec. 15.3.3). The appearance of the second knee may also be due to the limitations of galactic cosmic-ray sources, but in this case corresponding to the maximum energy to which iron nuclei are accelerated. As the contributions from galactic sources become suppressed at energies above the knee and second knee, extragalactic sources capable of accelerating particles to higher energies begin to dominate the cosmic-ray spectrum, possibly leading to the appearance of the ankle [267, 268]. The final downturn of this spectrum at energies above $\sim 3 \times 10^{10}$ GeV is, at least in part, due to the scattering of protons with the CMB (and nuclei with the infrared background), giving rise to the spectral feature known as the Greisen-Zatsepin-Kuzmin (GZK) cutoff [269, 270] (see Sec. 15.4).

15.2 Cosmic-ray propagation

The magnetic fields that are encountered by cosmic rays are often irregular, leading to motion that can treated as a random walk. This behavior can be quantified in terms of a scalar diffusion coefficient, $D(\vec{x})$. Through diffusion, the number density of cosmic rays,

$n(\vec{x})$, evolves according to the following differential equation:

$$\frac{dn(\vec{x})}{dt} = D(\vec{x})\,\nabla^2 n(\vec{x}) \tag{15.1}$$
$$= D(\vec{x})\left[\frac{\partial^2 n(\vec{x})}{\partial x^2} + \frac{\partial^2 n(\vec{x})}{\partial y^2} + \frac{\partial^2 n(\vec{x})}{\partial z^2}\right],$$

where the diffusion coefficient has units of area per time. Note that this expression describes only the effects of diffusion, and does not account for any energy loss processes or for the creation or destruction of particles.

Over the course of a random walk in a region with a uniform diffusion coefficient D, a typical particle will become displaced after a time t by a distance of $\ell \sim L_{\rm dif} \equiv 2\sqrt{Dt}$. For a population of particles that were injected from a given source at a time $t = 0$, the solution to Eq. 15.1 is

$$n \propto e^{-d^2/L_{\rm dif}^2}, \tag{15.2}$$

where d is the distance to the source. Very roughly speaking, these particles uniformly fill a spherical volume of radius $\sim L_{\rm dif}$ around the source.

Including the effects of energy losses, Eq. 15.1 becomes [25]

$$\frac{dn(\vec{x},E)}{dt} = D(\vec{x},E)\,\nabla^2 n(\vec{x},E) + \frac{d}{dE}\left[b(\vec{x},E),\ n(\vec{x},E)\right] + Q(\vec{x},E), \tag{15.3}$$

where $n(\vec{x},E)$ is now the number density of cosmic rays per unit energy, $b(\vec{x},E) = -dE/dt$ is the energy loss rate of the cosmic rays, and $Q(\vec{x},E)$ is the rate at which new cosmic rays are created (or accelerated) per unit time, energy, and volume. This diffusion-loss equation is often solved under the assumption that the system is in a steady-state configuration ($dn/dt = 0$). If this is not the case, then n and Q should be treated as functions of time.

In some cases, cosmic rays can be created or destroyed through scattering or decays. For example, we can write down the following set of diffusion-loss equations to describe the system of cosmic rays in the form of boron and carbon nuclei:

$$\frac{dn_{\rm C}(\vec{x},E_{\rm C})}{dt} = D(\vec{x},E_{\rm C})\,\nabla^2 n_{\rm C}(\vec{x},E_{\rm C}) - \frac{n_{\rm C}(\vec{x},E_{\rm C})}{\tau_{\rm C}(\vec{x},E_{\rm C})} + Q_{\rm C}(\vec{x},E_{\rm C}), \tag{15.4}$$

$$\frac{dn_{\rm B}(\vec{x},E_{\rm B})}{dt} = D(\vec{x},E_{\rm B})\,\nabla^2 n_B(\vec{x},E_{\rm B}) + \frac{n_{\rm C}(\vec{x},E_{\rm C})}{\tau_{\rm C}(\vec{x},E_{\rm C})}\,BR({\rm C}\rightarrow{\rm B}),$$

where $\tau_{\rm C}(\vec{x},E_{\rm C})$ is the timescale for a carbon nucleus to be destroyed via spallation, and $BR({\rm C}\rightarrow{\rm B})$ is the fraction of those collisions that result in the production of a boron nucleus. The spallation timescale can be written as

$$\tau_{\rm C}(\vec{x},E_C) = \left[\sum_i \sigma_i(E_C)\,n_i(\vec{x})\,v(E_C)\right]^{-1}, \tag{15.5}$$

where σ_i is the cross section to break up a carbon nucleus in a collision with gas of species i. The values of $E_{\rm C}$ and $E_{\rm B}$ are related in these equations such that the carbon and boron nuclei

each have the same kinetic energy per nucleon. Note that I have included a term in these equations for the creation of carbon, Q_C, but not for boron (except through the spallation of carbon), as the sources of the cosmic rays do not accelerate significant quantities of boron.

Boron is an example of a stable secondary species, which is produced in large part through the spallation of carbon nuclei. Stable secondary-to-primary ratios, such as boron-to-carbon, can be used to constrain the column density of gas that is encountered over the lifetime of a typical cosmic ray. In addition, observations of unstable secondaries, such as beryllium-10, can be used to constrain the amount of time that cosmic rays propagate. By combining measurements of stable and unstable secondary-to-primary ratios as a function of kinetic energy per nucleon, one can obtain a reasonable determination of the diffusion coefficient and its energy dependance, as well as of the size of the region through which the cosmic rays diffuse [25].

The energy dependence of the diffusion coefficient is often parameterized as follows:

$$D(R) = D_0\, v \left(\frac{R}{R_0}\right)^{\alpha}, \tag{15.6}$$

where v is the velocity of a particle, and R is its rigidity, defined as momentum per unit charge.

Measurements of the cosmic-ray spectrum at $\sim$GeV–TeV energies indicate that these particles propagate in the vicinity of the galactic disk with a diffusion coefficient that is approximately described by $D_0 \sim 4 \times 10^{28}\,\mathrm{cm}^2/\mathrm{s}$ and $\alpha \sim 0.3$, for $R_0 \equiv 1\,\mathrm{GV}$ (equivalent to $p/q = 1\,\mathrm{GeV}$). This diffusion coefficient appears to provide a reasonable description for the propagation of cosmic rays throughout much the galactic disk and the surrounding volume. Several kiloparsecs from the galactic disk, the conditions transition to a state in which the cosmic rays are instead effectively free streaming, traveling with less deflection and along increasingly straight trajectories.

Combining the diffusion length with the energy loss rates from synchrotron and inverse Compton processes, as described in Chapter 5, we can estimate the distance that a typical cosmic-ray electron will travel from its origin before it loses most of its energy. In the Thomson limit, this distance is given by

$$\begin{aligned} L_{\rm dif}(E_e) &= 2\left[D(E_e)\left(\frac{dE_e}{dt}\frac{1}{E_e}\right)^{-1}\right]^{1/2} \\ &\approx \left[\frac{3D_0 m_e^2}{\sigma_t(\rho_{\rm rad}+\rho_B)E_e^{1-\alpha}}\right]^{1/2} \\ &\approx 10\,\mathrm{kpc} \times \left(\frac{D_0}{4\times10^{28}\,\mathrm{cm}^2/\mathrm{s}}\right)^{1/2} \left(\frac{1.6\,\mathrm{eV/cm}^3}{\rho_{\rm rad}+\rho_B}\right)^{1/2} \left(\frac{\mathrm{GeV}}{E_e}\right)^{(1-\alpha)/2}, \end{aligned} \tag{15.7}$$

where in the second line I have assumed that the electrons are relativistic, $E_e \gg m_e$. From this expression, we see that while a GeV-scale electron can propagate across much of the galaxy before losing most of its energy, a TeV-scale electron will cool much more rapidly, losing a significant fraction of its energy before diffusing even a kiloparsec from its source.

15.3 Cosmic-ray acceleration

15.3.1 Second order fermi acceleration

In 1949, Enrico Fermi proposed a mechanism for the acceleration of cosmic rays which, for reasons we will see shortly, has become known as "second order Fermi acceleration" [271]. In this mechanism, cosmic rays are deflected by massive clouds, which act as magnetic mirrors. While such deflections leave the energy of a cosmic ray unchanged in the center-of-momentum frame of the collision, the energy of the particle in the lab frame changes by the following degree (up to second order in v/c):

$$\frac{\Delta E}{E} \approx 2 v_{\rm cloud} v_{\rm CR} \cos\theta + 2 v_{\rm cloud}^2, \tag{15.8}$$

where $v_{\rm cloud}$ and $v_{\rm CR}$ are the speeds of the cloud and the cosmic ray, each in units of the speed of light, and θ is the angle of the cosmic-ray's trajectory, relative to a head-on collision with the moving cloud. In collisions with $\cos\theta > -v_{\rm cloud}/v_{\rm CR}$, the incoming particle will gain energy. Otherwise, it will lose energy in the interaction.

What we are interested in here is how much energy a cosmic ray gains or loses in an *average* collision. For an isotropic distribution of clouds and particles, the probability that a collision will occur at a given angle, θ, is proportional to $P(\theta) \propto (1 + v_{\rm cloud} \cos\theta)$, allowing us to calculate the average change in energy:

$$\begin{aligned}
\left\langle \frac{\Delta E}{E} \right\rangle &\approx \frac{\int_0^\pi (2 v_{\rm cloud} v_{\rm CR} \cos\theta + 2 v_{\rm cloud}^2)\,(1 + v_{\rm cloud}\cos\theta)\,\sin\theta d\theta}{\int_0^\pi (1 + v_{\rm cloud}\cos\theta)\,\sin\theta d\theta} \\
&= \frac{\int_0^\pi (2 v_{\rm cloud} v_{\rm CR} \cos\theta + 2 v_{\rm cloud}^2 + 2 v_{\rm cloud}^2 v_{\rm CR} \cos^2\theta + 2 v_{\rm cloud}^3 \cos\theta)\,\sin\theta d\theta}{\int_0^\pi (1 + v_{\rm cloud}\cos\theta)\,\sin\theta d\theta} \\
&= 2 v_{\rm cloud}^2 + \frac{2 v_{\rm cloud}^2 v_{\rm CR}}{3} \\
&\approx \frac{8 v_{\rm cloud}^2}{3},
\end{aligned} \tag{15.9}$$

where in the last step I have taken the cosmic ray to be relativistic, $v_{\rm CR} \approx 1$. It is because this average change in energy depends on the square of the cloud's velocity, $v_{\rm cloud}^2$, that this mechanism is known as second order Fermi acceleration.

Since the result shown in Eq. 15.9 is positive, we expect the cosmic rays to increase in energy, at the expense of the kinetic energy of the clouds. For a scattering rate, R, we predict

$$\begin{aligned}
\frac{dE}{dt} &\approx R \left\langle \frac{\Delta E}{E} \right\rangle E \\
&\approx \frac{8 v_{\rm cloud}^2 R E}{3},
\end{aligned} \tag{15.10}$$

whose solution is $E(t) = E_i\, e^{8 v_{\rm cloud}^2 R t/3}$. Thus, under these assumptions, the energies of the particles should increase exponentially with time.

Realistically, the accelerating particles will remain within the acceleration region (which contains the clouds) for only a finite period of time. If we model their escape according to $dN/dt = -N/\tau_{\rm esc}$, where $\tau_{\rm esc}$ is the timescale for escape from the region, we obtain the following steady state solution for the particles that remain within the region:

$$\begin{aligned} 0 = \frac{dN(E)}{dt} &= \frac{d}{dE}\left[-\frac{dE}{dt} N(E) \right] - \frac{N(E)}{\tau_{\rm esc}} \\ &= \frac{\partial}{\partial E}\left[-\frac{8 v_{\rm cloud}^2 R E}{3} N(E) \right] - \frac{N(E)}{\tau_{\rm esc}} \\ &= -\frac{8 v_{\rm cloud}^2 R}{3} N(E) - \frac{dN(E)}{dE} \frac{8 v_{\rm cloud}^2 R E}{3} - \frac{N(E)}{\tau_{\rm esc}} \\ \longrightarrow \quad \frac{dN(E)}{dE} &= -\frac{N(E)}{E}\left[1 + \frac{3}{8 v_{\rm cloud}^2 R\, \tau_{\rm esc}} \right], \end{aligned} \tag{15.11}$$

where $N(E)$ is the number of cosmic rays in the region per unit energy. The solution to this differential equation is a power-law spectrum:

$$N(E) \propto E^{-1-\left(3/8 v_{\rm cloud}^2 R\, \tau_{\rm esc}\right)}. \tag{15.12}$$

Fermi's proposed mechanism can indeed accelerate particles, leading to a power-law spectrum of cosmic rays, at least under certain idealized conditions. At this point, however, we can ask ourselves whether this is a realistic mechanism to explain the observed cosmic-ray spectrum. The answer, sadly, is no. Most importantly, under realistic astrophysical circumstances, this process will be extremely slow. For $v_{\rm cloud} \sim 10^{-4}$ and $R \sim c/(0.1\,{\rm pc})$, for example, appreciable acceleration will occur only on timescales of $\tau \equiv E\,(dE/dt)^{-1} \sim 4 \times 10^{14}\,{\rm s} \approx 12\,{\rm Myr}$. Furthermore, energy loss processes are often competitive with those of acceleration, preventing this mechanism from even slowly increasing the energies of cosmic rays.

15.3.2 First order fermi acceleration

In Fermi's original proposal for cosmic-ray acceleration, the collisions are taken to be isotropically oriented. This leads to acceleration that is second order in velocity, severely limiting the efficiency of this mechanism. In some astrophysical environments, however, the geometry can be more favorable for the purposes of cosmic-ray acceleration. In particular, acceleration can be much more efficient in situations in which the particles are preferentially moving along the direction of their targets' motion. This can be the case, for example, in environments with strong (i.e., highly supersonic) shock waves that propagate outward in spherical shells from their source. Since the 1970s, this mechanism of diffusive shock acceleration has been the dominant idea for how charged particles are accelerated in astrophysical environments [272, 273, 274].

In deriving Eq. 15.9, I adopted an isotropic distribution for the velocities of the particles and the clouds, corresponding to a distribution of scattering angles that is given by $P(\theta) \propto (1 + v_{\rm cloud} \cos\theta)$. In contrast, in diffusive shock acceleration, the shock waves move radially outward from their source, while the particles in question move inward or outward along approximately radial trajectories. In such a scenario, the collisions tend to be approximately

head-on, causing each collision to increase the particle's energy by a quantity that is proportional to the velocity of the cloud, $\Delta E/E \approx 2 v_{\rm cloud}\, v_{\rm CR}$ (see Eq. 15.8).

With this in mind, let's consider an example in which the distribution of scattering angles is given by $P(\theta) \propto \cos\theta$, somewhat favoring collisions that are approximately head-on. For this distribution, the average shock crossing leads to the following change in the energy of a charged particle:

$$\left\langle \frac{\Delta E}{E} \right\rangle \approx \frac{\int_0^{\pi/2} (2 v_{\rm shock}\, v_{\rm CR} \cos\theta + 2 v_{\rm shock}^2)\ \sin\theta \cos\theta\, d\theta}{\int_0^{\pi/2} \sin\theta \cos\theta\, d\theta} \tag{15.13}$$

$$= \frac{4 v_{\rm shock}\, v_{\rm CR}}{3} + 2 v_{\rm shock}^2$$

$$\approx \frac{4 v_{\rm shock}}{3}.$$

As promised, this differs from the result found in the previous subsection in that it is first order in velocity, allowing us to accelerate cosmic rays much more efficiently.

To place first order Fermi acceleration within the context of a more concrete scenario, we can consider the shock waves that move through the interstellar medium ahead of the supersonic shells of a supernova remnant. In this environment, cosmic rays are present on either side of a given shock, and are moving much faster than the shock itself. Each collision impacts a given particle only mildly, and the particles ultimately move across the shocks many times in both directions, statistically gaining energy with each crossing.

Each time a cosmic ray collides with a shock, there is a probability given by $P \approx 1 - u$ that the particle will remain within the acceleration region, where u is the velocity of the shock relative to the medium (such as the surrounding gas). Repeating the calculation described in Eq. 15.11, but using $dE/dt = (4 v_{\rm shock} R/3)\, E$ and $\tau_{\rm esc} = 1/R\,(1-P) \approx 1/Ru$, we find that first order Fermi acceleration should yield a spectrum of particles that is given by $N(E) \propto E^{-1-3u/4v_{\rm shock}}$.

We can relate the values of u and $v_{\rm shock}$ by considering the environments on either side of a given shock. By conserving the mass that crosses a shock, we can write $\rho_1\, u = \rho_2\, v_{\rm shock}$, where ρ_1 and ρ_2 are the densities of gas immediately ahead of and behind the shock, as evaluated in the rest frame of the shock. In the case of a strong shock, we can further relate $\rho_2/\rho_1 = (1+r)/(1-r)$, where r is the ratio of the specific heat capacities of the gas. For the case of a monatomic or ionized gas, this ratio is given by $r = 5/3$, and thus $\rho_2/\rho_1 = 4$ and $u/v_2 = 1/4$. Moving into the rest frame of the gas, this implies that $v_{\rm shock} = (3/4)\, u$. In other words, while the shock is moving through the gas at a velocity u, the gas behind the shock is moving at a somewhat slower speed of $3u/4$ relative to the upstream gas. When this is combined with the result found in the previous paragraph, we find that first order Fermi acceleration should produce cosmic rays with a spectrum of the form $N(E) \propto E^{-1-3u/4v_{\rm shock}} \approx E^{-2}$, at least under these idealized conditions.

Strong shocks of the kind that can lead to first order Fermi acceleration are found across a wide range of astrophysical environments, including supernova remnants, active galactic nuclei, and gamma-ray bursts, among others. In practice, real-world cosmic-ray accelerators do not always produce particles with a spectrum of the form $dN/dE \propto E^{-2}$. After all, we ignored the effects of energy losses in that calculation, and otherwise worked under a number of idealized assumptions. That being said, diffusive shock acceleration is rather

generically predicted to produce cosmic rays with a roughly power-law spectrum, typically with a spectral index of around −2.

15.3.3 The maximum energy of cosmic-ray acceleration

Although diffusive shock acceleration can produce highly energetic cosmic rays, this mechanism has its limitations. In particular, the maximum energy to which a given particle can be accelerated depends on the geometry of the acceleration region, and on the strength of the magnetic field that is present.

In the process of first order Fermi acceleration, it follows from Eq. 15.13 that a particle with an initial energy E_0 will be accelerated after k scattering events (i.e., shock crossings) to the following energy:

$$E \approx E_0 \left(1 + \frac{4 v_{\rm shock}}{3}\right)^k. \tag{15.14}$$

For the case of a young supernova remnant with $v_{\rm shock} \sim 0.1\,c$, for example, it would take $k \sim 10^2$ crossings for a particle with an initial energy of $E_0 \sim 1\,{\rm GeV}$ to reach an energy of $E \sim 1\,{\rm PeV}$. By this time, many of the cosmic rays will have already escaped the acceleration region, preventing them from reaching arbitrarily high energies.

We can apply similar logic to construct a more general version of this argument, applicable to a wide range of astrophysical sources and environments. From Faraday's law, $\nabla \times \vec{E} = -\partial \vec{B}/\partial t$, we can roughly relate the strengths of the electric and magnetic fields that are present within a region:

$$\frac{E}{L} \sim \frac{B}{L/u}, \tag{15.15}$$

where L is the size of the region and u is the velocity of the shocks. From the strength of the electric field, $E \sim Bu$, we can integrate over a path to estimate the maximum energy to which a charged particle can be accelerated:

$$\begin{aligned} E_{\rm max} &= Z \int E\, dx \\ &\sim ZBuL \\ &\approx 10^6\,{\rm GeV} \times \left(\frac{Z}{1}\right)\left(\frac{B}{\mu{\rm G}}\right)\left(\frac{u}{0.1\,c}\right)\left(\frac{L}{10\,{\rm pc}}\right), \end{aligned} \tag{15.16}$$

where Z is the charge of the particle being accelerated, and E in the integrand is the strength of the electric field. Thus, for a young supernova remnant with $B \sim \mu{\rm G}$, $u \sim 0.03\,c$, and $L \sim 0.03\,c \times 10^3\,{\rm yrs} \sim 9.2\,{\rm pc}$, this expression yields a maximum energy of about $E_{\rm max} \sim 300\,{\rm TeV} \times Z$. From this and other arguments, it appears that supernova remnants should be able to accelerate cosmic rays only up to PeV-scale energies. At significantly higher energies, other kinds of sources must be responsible for the observed cosmic-ray spectrum.

15.4 Ultra-high energy cosmic rays

The measured cosmic-ray spectrum extends up to energies as high as a few times $10^{11}\,{\rm GeV}$ [267, 275]. At these highest energies, the cosmic rays consist of a mixture of nuclear species, on average heavier than protons but lighter than iron [276, 277, 278]. From Eq. 15.16, it is clear that these particles must originate in exceptional environments, containing strong magnetic fields that are extended over large regions of space,

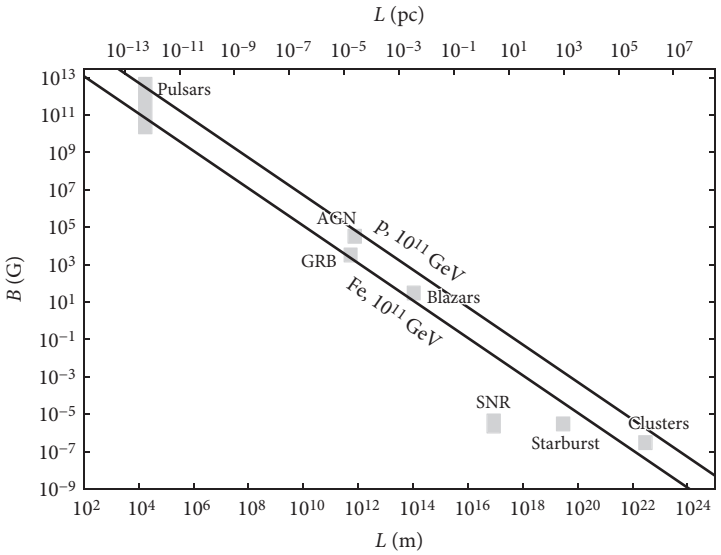


Figure 15.2 A "Hillas plot" [279], showing the magnetic field strength and the size of the acceleration region that are required for a source to potentially accelerate protons or iron nuclei to $E \geq 10^{11}$ GeV. Also shown are typical parameter values associated with various potential sources of high- and ultra-high energy cosmic rays.

$(B/\mathrm{G})(L/\mathrm{pc}) \gtrsim 3 \times 10^{10}\, Z^{-1}$. In the context of ultra-high energy cosmic rays, these requirements are often expressed in terms of what is called a "Hillas plot" [279]. An example of a Hillas plot is shown in Figure 15.2, indicating the values of the magnetic field strength and the size of the acceleration region that would be required in order for a source to potentially accelerate either protons or iron nuclei to an energy of $E = 10^{11}$ GeV. Also shown in this plane are typical parameter values associated with various potential sources of high and ultra-high energy cosmic rays, including pulsars, active galactic nuclei (AGN), gamma-ray bursts (GRB), blazars, starburst winds, and galaxy cluster shocks. For comparison, I've also included supernova remnants (SNR) in this plot, which cannot accelerate particles to ultra-high energies. One should keep in mind that these parameter values are only intended to reflect representive estimates, and that no two sources of cosmic rays are exactly alike.

The approximate isotropy of the highest energy cosmic rays suggests that these particles originate from a large number of extragalactic sources. In particular, whereas lower energy cosmic rays are efficiently isotropized by the magnetic field of the Milky Way, this will not be the case for the highest energy particles. To see why, consider a charged particle as it travels through a random walk across a series of regions, each with its own magnetic field orientation. As the particle travels through one of these regions, it will be deflected by an angle given by $\sim L_{\rm coh}/R_L$, where $L_{\rm coh}$ is the size (or coherence length) of the magnetic field domain, and $R_L = \gamma m v_\perp / ZeB$ is the Larmour radius of the particle. After traveling a distance d, and thus traveling across $d/L_{\rm coh}$ domains, the particle will have experienced approximately the following degree of net deflection:

$$\theta \sim \left(\frac{L_{\rm coh}}{R_L}\right)\left(\frac{d}{L_{\rm coh}}\right)^{1/2} \tag{15.17}$$
$$\sim 0.36^\circ \times \left(\frac{10^{11}\,\mathrm{GeV}}{E}\right)\left(\frac{d}{1\,\mathrm{kpc}}\right)^{1/2}\left(\frac{L_{\rm coh}}{50\,\mathrm{pc}}\right)^{1/2}\left(\frac{B}{3\,\mu\mathrm{G}}\right)\left(\frac{Z}{1}\right).$$

While this magnitude of deflection is not negligible (especially for nuclei with $Z \gg 1$), it is also not sufficient to effectively isotropize the arrival directions of the highest energy cosmic rays, especially if these particles originate from a small number of sources or from a population of galactic sources. In contrast, we expect cosmic-ray protons with energies below roughly $\sim 10^8 - 10^9$ GeV to be throughly isotropized by the magnetic field of our galaxy.

If we repeat this calculation for the case of cosmic rays propagating through the intergalactic magnetic field, we reach a similar conclusion:

$$\theta \sim 0.17^\circ \times \left(\frac{10^{11}\,\text{GeV}}{E}\right)\left(\frac{d}{10\,\text{Mpc}}\right)^{1/2}\left(\frac{L_{\text{coh}}}{1\,\text{Mpc}}\right)^{1/2}\left(\frac{B}{0.1\,\text{nG}}\right)\left(\frac{Z}{1}\right). \tag{15.18}$$

In this case, however, it should be noted that the properties of the intergalactic magnetic field are highly uncertain. In particular, the magnetic fields found throughout the intergalactic medium could be significantly weaker than 0.1 nG.

Although the sources of the highest energy cosmic rays are very likely extragalactic in nature, they cannot be arbitrarily far away. In particular, the scattering of ultra-high energy protons and nuclei with background radiation limits the distance from which these particles can reach us, and leads to the spectral feature known as the GZK cutoff [269, 270]. In the case of protons, this occurs through their scattering with the CMB:

$$\begin{aligned} p + \gamma_{\text{CMB}} &\longrightarrow n + \pi^+ \longrightarrow p\, e^- \bar{\nu}_e + e^+ \nu_e \nu_\mu \bar{\nu}_\mu, \\ p + \gamma_{\text{CMB}} &\longrightarrow p + \pi^0 \longrightarrow p + \gamma\gamma. \end{aligned} \tag{15.19}$$

Due to the large cross sections involved (on the order of $\sigma \sim 10^{-27}$ cm^2; see Chapter 5), and the large number density of CMB photons, these scattering processes can be very efficient. For them to take place, however, the energy in the center-of-momentum frame must be large enough to exceed the threshold for pion production:

$$E_{\text{CM}} = \sqrt{2 E_p E_\gamma} > m_p + m_\pi. \tag{15.20}$$

For $E_\gamma \sim 2.7\, T_{\text{CMB}}$, this requirement corresponds to $E_p \gtrsim 10^{12}$ GeV. Taking into account the width of the CMB's blackbody energy distribution, this scattering process is quite efficient down to energies of a few times 10^{10} GeV. Above the energy threshold for these processes, the mean free path is roughly given by

$$\begin{aligned} \lambda(E_p) &= \left[\int \sigma(E_p, E_\gamma)\, \frac{dN_\gamma}{dE_\gamma}(E_\gamma)\, dE_\gamma\right]^{-1} \\ &\sim 30\,\text{Mpc}. \end{aligned} \tag{15.21}$$

Ultra-high energy protons also experience significant energy losses from pair production, $p + \gamma_{\text{CMB}} \longrightarrow p\, e^+ e^-$ [280]. While the energy loss rate from this process is less dramatic than that from pion production, it is effective at significantly lower energies, providing the dominant energy loss mechanism for protons with energies between $E \sim 3 \times 10^9$ and $\sim 5 \times 10^{10}$ GeV.

In the case of ultra-high energy nuclei, scattering with the infrared and microwave backgrounds leads to their efficient photo-disassociation, breaking apart most nuclear species at the highest observed energies on length scales as short as $\lambda \sim 1 - 10$ Mpc [268, 281].

This has the effect of attenuating the spectrum of cosmic-ray nuclei, as well as altering its composition from heavier nuclei toward lighter nuclear species, including protons.

The downturn that is observed in the cosmic-ray spectrum at energies above a few times 10^{10} GeV [264, 263, 282, 265] is consistent with expectations associated with these scattering processes. Furthermore, these considerations indicate that the highest energy cosmic rays must predominantly originate from sources that reside within roughly ~ 100 Mpc of the Milky Way. Recently reported evidence of a modest anisotropy in the arrival directions of the ultra-high energy cosmic rays [283, 264, 284] may provide us with some clues regarding the origin of these particles. That being said, it remains difficult to disentangle this information from the impact of the galactic magnetic field on the trajectories of the cosmic rays.

Additional reading

When teaching graduate-level courses on high-energy astrophysics, I have often used the textbook by Longair. I also recommend the Particle Data Group's very useful review of cosmic rays [24].

Problems

15.1 Consider an object that produces 10^{40} cosmic-ray electrons per second, each with an energy of 1 TeV. For a diffusion coefficient of $D = 10^{28}$ cm^2/s, and an energy loss rate of $dE_e/dt = 10^{-16}$ GeV/s $\times$ $(E_e/\text{GeV})^2$, estimate the flux of these electrons with $E_e \gtrsim 100$ GeV at a distance of 1 kpc from the source.

15.2 The Geminga pulsar is surrounded by a region of radius ~ 10 pc that emits a large flux of $\sim$ 10–50 TeV gamma rays through the process of inverse Compton scattering. What does this information tell us about the diffusion coefficient in the region surrounding Geminga? (Hint: See Ref. [285].)

15.3 Derive the expression given in Eq. 15.8. Hint: See Sec. 17.3 of Longair's *High-Energy Astrophysics*.

15.4 Repeat the calculation described in Sec. 15.3.2 for a scenario in which every collision occurs exactly head-on, $\cos\theta = 1$. What is the cosmic-ray spectrum that results in this case?

15.5 Apply Eq. 15.16 to the case of the Large Hadron Collider, using $B_{\rm LHC} \approx 8.3$ T and $L \sim R_{\rm LHC} \approx 4.3$ km. How does this result compare to the actual maximum energy to which protons are accelerated by this machine?

15.6 Imagine a scenario in which half of the magnetic domains in our galaxy have $B = 3\,\mu$G, and the other half have $B = 30\,\mu$G. How much deflection would we expect a cosmic-ray proton to experience as a function of energy? Adopt $L_{\rm coh} = 50$ pc.

16

Gamma-Ray and Neutrino Astrophysics

The interactions of cosmic rays can lead to the production of high-energy photons and neutrinos. Efforts to detect and study these particles have become a major thrust of modern astrophysics. In this chapter, I describe the processes that most readily produce these particles, including both hadronic and leptonic mechanisms. I also discuss the interactions experienced by very high-energy photons as they travel across cosmological distances, as well as the oscillations of astrophysical neutrinos. Finally, I review some of the astrophysical sources that are often discussed within the context of gamma-ray and neutrino astrophysics, including active galactic nuclei, gamma-ray bursts, supernova remnants, and pulsars. The information presented in this chapter will be important to our discussion of indirect searches for dark matter in Chapter 19.

16.1 Why gamma rays? Why neutrinos?

Prior to the middle of the twentieth century, astronomers relied almost entirely on photons with visible wavelengths to study our universe. In the 1950s, however, scientists began to observe the sky using telescopes that were capable of detecting light in the radio, infrared, and X-ray bands. Shortly thereafter, in the 1960s, the first gamma-ray telescopes were deplod. Light at different wavelengths carries different information, and that information can be used to answer different questions. Gamma rays, in particular, serve as an essential probe of some of our universe's most extreme and high-energy environments.

Astrophysical neutrinos also provide unique advantages for astronomers. In particular, neutrinos are rarely absorbed or attenuated, even across cosmological distances, or through dense environments. Neutrinos can thus probe regions that are inaccessible by other means. Furthermore, high-energy neutrinos are predominantly produced through interactions involving nucleons or nuclei, and thus provide us with a powerful way of determining whether a given object, or class of objects, is capable of accelerating cosmic rays. The interplay between gamma-ray astronomy, neutrino astronomy, and cosmic-ray physics has become a major theme of modern particle astrophysics.

16.2 The astrophysical production and propagation of gamma rays

16.2.1 Gamma rays from leptonic processes

Let's begin by considering astrophysical objects and environments that produce gamma-ray emission through processes involving high-energy electrons and positrons (i.e., bremsstrahlung and inverse Compton scattering). Such objects are known as leptonic sources. As described in Sec. 5.1.1, bremsstrahlung emission is generated when energetic electrons pass through the electric fields of charged particles, such as those found in clouds of gas. Inverse Compton emission is produced when an energetic electron collides with a low-energy photon, increasing its energy into the X-ray or gamma-ray range.[1]

As described in Sec. 15.3.2, first order Fermi acceleration generally leads to the production of a spectrum of charged particles with an approximately power-law form, $dN_e/dE_e \propto E_e^{-\gamma}$, where γ is typically around $\sim$2. Unlike hadrons and nuclei, cosmic-ray electrons lose energy fairly rapidly, especially at high energies, limiting the maximum energy to which such particles can be accelerated.

The characteristics of the gamma-ray emission from a leptonic source depend not only on the intensity and spectrum of the accelerated electrons, but also on the density and spectrum of the surrounding radiation fields, magnetic fields, and gas. These many factors lead to a diversity of gamma-ray spectra from accelerators of cosmic-ray electrons. That being said, a few features are commonly found across a wide range of such objects.

At sufficiently high energies ($E_e \gg m_e^2/\epsilon_\gamma$), electrons undergo Inverse Compton scattering in the Klein-Nishina regime, in which each upscattered photon carries an energy that is approximately equal to that of the incoming electron, $E_\gamma \sim E_e$ (see Sec. 5.1.3). In this energy range, we can treat $d\sigma_{\rm ICS}/dE_\gamma \propto \delta(E_e - E_\gamma)$, leading to the following spectral shape for the resulting gamma-ray emission:

$$\frac{dN_\gamma}{dE_\gamma}(E_\gamma) = \int\int \frac{dN_e}{dE_e}(E_e)\,\frac{d\sigma_{\rm ICS}}{dE_\gamma}(E_\gamma,\epsilon_\gamma,E_e)\,\frac{dn_\gamma}{d\epsilon_\gamma}(\epsilon_\gamma)\,dE_e\,d\epsilon_\gamma \tag{16.1}$$

$$\approx \sigma_{\rm ICS}\,n_\gamma \int \frac{dN_e}{dE_e}(E_e)\,\delta(E_e - E_\gamma)\,dE_e,$$

where $dn_\gamma/d\epsilon_\gamma$ is the spectrum of the target radiation and $\sigma_{\rm ICS}\,n_\gamma = \int (d\sigma_{\rm ICS}/dE_\gamma)\,(dn_\gamma/d\epsilon_\gamma)\,d\epsilon_\gamma$. The gamma-ray spectrum produced in the Klein-Nishina regime thus traces that of the scattering electrons.

At lower energies, in the Thomson regime, the resulting gamma rays carry significantly less energy than the incoming electron. More specifically, the differential cross section for inverse Compton scattering is approximately flat with $d\sigma_{\rm ICS}/dE_\gamma \approx 3\sigma_t m_e^2/4\epsilon_\gamma$ up to a maximum energy of $E_\gamma \approx 4\epsilon_\gamma E_e^2/m_e^2$, and thus the typical gamma ray carries only a small fraction of the energy of the scattering electron, $E_\gamma/E_e \sim 2E_e\epsilon_\gamma/m_e^2 \ll 1$. This allows us to estimate the resulting gamma-ray spectrum as follows:

$$\frac{dN_\gamma}{dE_\gamma}(E_\gamma) = \int\int \frac{dN_e}{dE_e}(E_e)\,\frac{d\sigma_{\rm ICS}}{dE_\gamma}(E_\gamma,\epsilon_\gamma,E_e)\,\frac{dn_\gamma}{d\epsilon_\gamma}(\epsilon_\gamma)\,dE_e\,d\epsilon_\gamma \tag{16.2}$$

[1] As I did in Chapter 15, I will often use the word "electron" in this discussion to refer to both electrons and positrons.

$$\approx \frac{3\sigma_t m_e^2 n_\gamma}{4\epsilon_\gamma} \int_0^{(E_\gamma m_e^2/4\epsilon_\gamma)^{1/2}} \frac{dN_e}{dE_e}(E_e)\, \frac{dE_e}{E_e^2}.$$

As an example, consider a population of electrons with a power-law spectrum, $dN_e/dN_e \propto E_e^{-\alpha}$. By carrying out the integral described above, we find that the gamma rays produced in the Thomson limit are also described by a power law, but with a spectral index of $-(\alpha+1)/2$.

For relatively low-energy electrons, in the Thomson regime, the energy-loss rates from the processes of synchrotron and inverse Compton scattering each scale as $dE_e/dt \propto E_e^2$, leading to a ratio that is determined simply by the energy densities of the magnetic and radiation fields:

$$\frac{(dE_e/dt)_{\rm syn}}{(dE_e/dt)_{\rm ICS}} = \frac{4\sigma_t \rho_{\rm mag} v_e^2 E_e^2/3m_e^2}{4\sigma_t \rho_{\rm rad} v_e^2 E_e^2/3m_e^2} = \frac{\rho_{\rm mag}}{\rho_{\rm rad}}. \tag{16.3}$$

At higher energies, where inverse Compton scattering takes place in the Klein-Nishina regime, the energy-loss rate from inverse Compton scattering is further suppressed by a factor of $\gamma_k^2/(\gamma_k^2+\gamma_e^2)$, where $\gamma_k = 3\sqrt{5}m_e/8\pi T$, T is the temperature of the target radiation, and $\gamma_e = E_e/m_e$. As a result, very high-energy electrons tend to lose a large fraction of their energy through synchrotron rather than inverse Compton scattering, suppressing the gamma-ray emission that is observed at the highest energies.

16.2.2 Gamma rays from hadronic processes

The cosmic-ray spectrum is not dominated by electrons, but rather by energetic protons and nuclei. When these particles are present in an astrophysical environment, they can produce pions through their collisions with gas or radiation. While charged pions decay into electrons and neutrinos, neutral pions decay into pairs of gamma rays.

As described in Sec. 5.2, the differential cross section for pion production in proton-proton collisions does not have a simple analytic form. For this reason, such calculations are often carried out using numerical tools which rely on tabulations of the relevant data [25, 26, 27, 28]. Roughly speaking, above the energy threshold for pion production, $E_p > E_{\rm th} \approx 1.2\,{\rm GeV}$, the total cross section for inelastic proton-proton (or antiproton-proton) scattering falls within the range of $\sigma_{pp} \sim (3\text{–}10)\times 10^{-26}\,{\rm cm}^2$. Furthermore, an average of approximately $N_\pi \approx 3\times[(E_p - E_{\rm th})/{\rm GeV}]^{1/4}$ pions are produced per collision, of which 1/3 are neutral and 2/3 are charged. Since higher energy protons create a greater multiplicity of pions per collision, the spectrum of the resulting pions (and thus the photons and neutrinos) is generally somewhat softer than that of the original cosmic-ray proton spectrum.[2] In particular, for a spectrum of high-energy cosmic-ray protons with a power-law form, $dN_p/dE_p \propto E_p^{-\Gamma_p}$, the resulting spectral index of the gamma rays and neutrinos will be $\Gamma_{\nu,\gamma} \approx (4/3)\Gamma_p - (2/3)$ (see Problem 16.1) [286, 287].

Pions can also be produced in the collisions of high-energy protons with photons, known as photohadronic interactions. Since the cross section for inelastic $p\gamma$ scattering is much smaller than that for pp scattering (compare Figs. 5.2 and 5.3), this process is generally only important in environments in which the number density of target photons is much greater than that of gas. The gamma-ray spectrum that results from these interactions depends both

[2] By "softer," I mean a spectrum that contains a larger proportion of low-energy particles, while a harder spectrum contains more particles of higher energy.

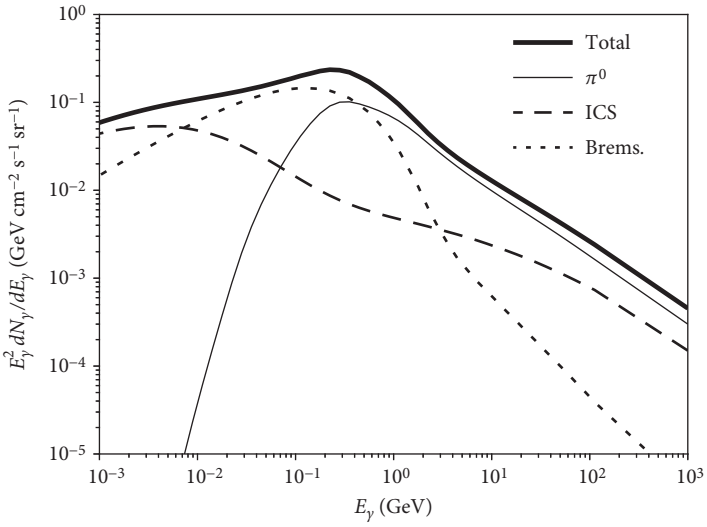


Figure 16.1 The spectrum of the diffuse gamma-ray emission from cosmic-ray interactions in the interstellar medium, averaged over the $40^\circ \times 40^\circ$ region centered around the Galactic Center, as calculated using Galprop [25].

on the spectrum of the incoming protons, as well as on the spectrum and intensity of the target radiation. Note that the cross section for $p\gamma$ scattering is significantly enhanced when the energy in the center-of-momentum frame is approximately equal to the mass of the delta baryon, $m_\Delta \approx 1.232$ GeV.

In Figure 16.1, I show the gamma-ray spectrum that is produced through the interactions of cosmic rays in the interstellar medium of our galaxy. This includes contributions from pion production, inverse Compton scattering, and bremsstrahlung (as calculated using the publicly available tool, Galprop [25]). The calculations behind these results include a model for the production and diffusion of cosmic rays, as well as for the distribution of the radiation fields, magnetic fields, and gas.

16.2.3 Gamma-ray propagation

At low energies, our modern universe is more or less transparent to gamma rays. At higher energies, however, gamma rays can scatter with low-energy photons, leading to the production of electron-positron pairs (see Sec. 5.3). Such electrons and positrons then lose their energy through inverse Compton scattering (as well as synchrotron), producing new gamma rays, and potentially generating a cascade of energetic photons, electrons, and positrons.

Consider a gamma ray that is emitted with an energy E_0 from a source at redshift z_s and that then travels through a background of photons with a differential number density given by $dn_\gamma/d\epsilon_\gamma$. From this information, we can calculate the resulting optical depth (see Eq. 5.32):

$$\tau(E_0, z_s) = \int_0^{z_s} dz \frac{dl(z)}{dz} \frac{1-\cos\theta}{2} d(\cos\theta) \int \sigma_{\gamma\gamma}(E_\gamma(z), \epsilon_\gamma, \theta) \frac{dn}{d\epsilon_\gamma}(\epsilon_\gamma, z)\, d\epsilon_\gamma, \quad (16.4)$$

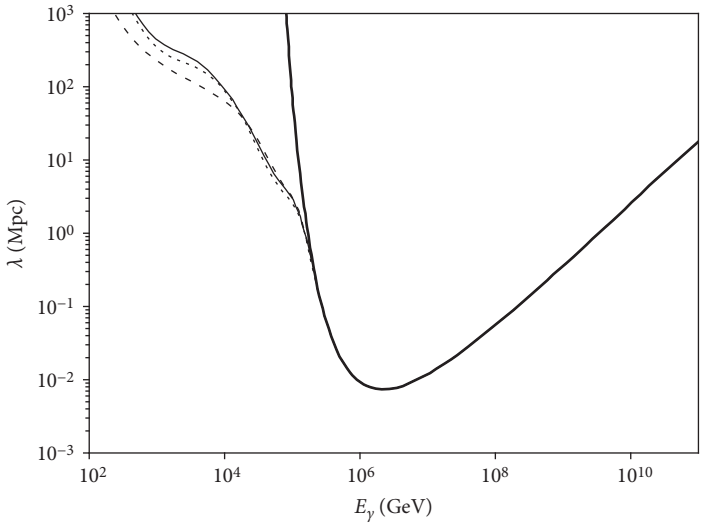


Figure 16.2 The mean free path for electron-positron pair production in the low-redshift universe. The thick solid line represents the result from scattering with the CMB, while the other curves include contributions from the cosmic infrared background (using three different models for this radiation [288, 289, 290]).

where $E_\gamma = E_0 \times (1+z)/(1+z_s)$, and $\sigma_{\gamma\gamma}$ is the cross section for pair production, as given in Eq. 5.30. The quantity dl/dz is the distance traveled by a photon per unit redshift:

$$\begin{aligned} \frac{dl}{dz} &= \frac{c}{H(z)(1+z)} \\ &= \frac{c}{H_0\left[\Omega_\Lambda + \Omega_M(1+z)^3\right]^{1/2}(1+z)}. \end{aligned} \tag{16.5}$$

Working in the $z_s \ll 1$ limit, we can relate the optical depth to the mean free path for pair production:

$$\lambda_\gamma(E_\gamma) = \frac{d_s}{\tau(E_\gamma, z_s)}, \tag{16.6}$$

where $d_s = cz_s/H_0$ is the distance to a nearby source at redshift z_s.

Due to scattering with the CMB, our universe is largely opaque to photons with $E_\gamma \gtrsim 10^5$ GeV, at least on cosmological distance scales. Furthermore, down to energies as low as $E_\gamma \sim \mathcal{O}(1\,\text{TeV})$, photons can produce electron-positron pairs by scattering with infrared photons (see Figure 16.2) [291, 292]. Once pair production has occurred, the newly formed electrons quickly lose their energy through inverse Compton scattering with the CMB. If the initial photon is sufficiently energetic, the resulting inverse Compton photons can have enough energy to undergo pair production themselves, leading to a recurring cycle of pair production and inverse Compton scattering, and resulting in the formation of an electromagnetic cascade. As an example, consider a photon with an energy of 10 PeV. Such a particle will quickly produce an electron and a positron, each of which will have an energy of $E_e \sim 5$ PeV. As this is safely within the Klein-Nishina regime ($E_e \gg m_e^2/\epsilon_{\text{CMB}}$), the subsequent inverse Compton scattering will produce new photons with energies of

approximately $E_\gamma \sim 5\,\text{PeV}$. This process continues until inverse Compton scattering no longer occurs in the Klein-Nishina regime, leading to the production of lower energy gamma rays that are below the threshold for pair production.

Through this process, any energy that is injected into the high-redshift universe in the form of $E_\gamma \gtrsim \text{TeV}$ photons gets transformed into an isotropic cascade of $\sim \text{GeV} - \text{TeV}$ gamma rays. For very high-energy gamma rays that are produced at cosmological distances, the resulting cascade has a spectrum that is approximately described by $dN_\gamma/dE_\gamma \propto E_\gamma^{-1.85}$ up to $E_\gamma \sim 10^2\,\text{GeV}$, above which the spectrum falls rapidly. The overall intensity of this cascade serves as a calorimeter, providing us with a measurement of the total amount of energy that has been injected into our universe in the form of very high-energy gamma rays, electrons, and positrons over cosmic history.

16.3 The astrophysical production and propagation of neutrinos

Unlike gamma rays, high-energy neutrinos are largely produced through hadronic interactions and, more specifically, through the decays of charged pions that are created in the collisions of high-energy protons. For this reason, the spectrum of high-energy neutrinos from a given source is expected to approximately trace that of the corresponding hadronic gamma-ray emission.[3]

Unlike very high-energy photons, neutrinos only rarely scatter and thus are not appreciably attenuated as they travel though our universe. They do, however, undergo oscillations, which impact the flavors of the neutrinos that reach earth. In the decays of charged pions ($\pi^+ \to e^+ \nu_e \bar{\nu}_\mu \nu_\mu, \pi^- \to e^- \bar{\nu}_e \nu_\mu \bar{\nu}_\mu$), neutrinos and antineutrinos are produced with a ratio of flavors given by $\phi_{\nu_e} : \phi_{\nu_\mu} : \phi_{\nu_\tau} = 1:2:0$. Since such particles traverse many oscillation lengths before reaching earth, we can safely apply the average, $\langle \sin^2(\Delta m^2 L/2E) \rangle \approx 0.5$, in evaluating Eq. 11.22. After making this substitution, the probability of flavor oscillations is given by

$$P_{\nu_\alpha \to \nu_\beta} = \delta_{\alpha\beta} - 2 \sum_{i>j} \text{Re}(U_{\alpha i} U^\star_{\beta i} U^\star_{\alpha j} U_{\beta j}) + \sum_{i>j} \text{Im}(U_{\alpha i} U^\star_{\beta i} U^\star_{\alpha j} U_{\beta j}), \tag{16.7}$$

where $U_{\alpha i}$ are elements of the neutrino mixing matrix (see Eqns. 11.14 and 11.15). Putting into this equation the appropriate entries of the PMNS matrix, as given in Eq. 11.15, we can calculate the oscillation probabilities (in vacuum). Making use of the fact that θ_{13} and δ_{CP} are each small, we can obtain a reasonably accurate, yet fairly simple, expression for these probabilities by taking these parameters to be zero. Under this approximation, we arrive at the following:

$$\begin{aligned} P_{\nu_e \to \nu_e} &\approx 1 - 2 s_{12}^2 c_{12}^2 \approx 0.57, \\ P_{\nu_e \to \nu_\mu} &= P_{\nu_\mu \to \nu_e} \approx 2 s_{12}^2 c_{23}^2 c_{12}^2 \approx 0.20, \\ P_{\nu_e \to \nu_\tau} &= P_{\nu_\tau \to \nu_e} \approx 2 s_{12}^2 s_{23}^2 c_{12}^2 \approx 0.23, \end{aligned} \tag{16.8}$$

[3] At lower energies, neutrinos can also be produced through fusion in stars, or through scattering processes that take place in dense environments, such as in the core of supernovae. Supernova 1987A, for example, produced a burst of MeV-scale neutrinos, of which 25 were detected on earth. In this chapter, we are instead going to focus on higher-energy neutrinos that are produced almost entirely through the decays of charged pions.

$$P_{\nu_\mu \to \nu_\mu} \approx 1 - 2(s_{12}^2 c_{12}^2 c_{23}^4 + s_{23}^2 c_{23}^2) \approx 0.41,$$

$$P_{\nu_\mu \to \nu_\tau} = P_{\nu_\tau \to \nu_\mu} \approx 2 s_{23}^2 c_{23}^2 (1 - s_{12}^2 c_{12}^2) \approx 0.39$$

$$P_{\nu_\tau \to \nu_\tau} \approx 1 - 2(s_{12}^2 c_{12}^2 s_{23}^4 + s_{23}^2 c_{23}^2) \approx 0.38,$$

where in the last step I have inserted the measured values of the mixing angles, as given in Eq. 11.27.

Combing this result with the initial flavor ratios predicted from pion decay, $\phi_{\nu_e} : \phi_{\nu_\mu} : \phi_{\nu_\tau} = 1 : 2 : 0$, we predict that the neutrinos will reach earth with the following flavor ratios [293, 294, 295, 296]:

$$\begin{aligned} \phi_{\nu_e} : \phi_{\nu_\mu} : \phi_{\nu_\tau} &= P_{\nu_e \to P_{\nu_e}} + 2P_{\nu_\mu \to P_{\nu_e}} : P_{\nu_e \to P_{\nu_\mu}} + 2P_{\nu_\mu \to P_{\nu_\mu}} : P_{\nu_e \to P_{\nu_\tau}} + 2P_{\nu_\mu \to P_{\nu_\tau}} \\ &\approx 1.07 : 1.02 : 0.99. \end{aligned} \tag{16.9}$$

In other words, over astrophysical distances, oscillations will transform the electron and muon neutrinos produced through pion decay into a roughly equal number of electron, muon, and tau neutrinos.

16.4 Astrophysical sources of high-energy gamma rays and neutrinos

In this section, I'll describe some of the kinds of astrophysical objects that are of most interest to gamma-ray and neutrino astronomers. While this will be far from a complete discussion of this subject, it should give the reader a flavor for the kinds of sources that are most often studied in this context.

16.4.1 The neutrino-cosmic ray connection

Before focusing on some specific sources of high-energy gamma rays and neutrinos, we're going to spend a couple of pages exploring the close connection that exists between the spectrum of high-energy cosmic rays and that of astrophysical neutrinos. Fundamentally, this connection arises from the simple fact that high-energy neutrinos are produced almost entirely through the interactions of high-energy protons.

In this discussion, we'll focus on the highest-energy cosmic rays, in the range of $E_{\rm CR} = 10^{10}$–10^{12} GeV. At these energies, the cosmic rays are extragalactic in origin. Furthermore, as a result of the GZK cutoff, the observed cosmic-ray spectrum is determined by the local ($\lesssim 100$ Mpc) injection rate (see Sec. 15.4). More specifically, the measured cosmic-ray spectrum in this energy range implies that cosmic rays are injected into the local universe at an average rate of

$$E_{\rm CR}^2 \frac{d\dot{N}_{\rm CR}}{dE_{\rm CR}}\bigg|_{z=0} \approx 6 \times 10^{46}\ {\rm GeV/Mpc^3/yr}. \tag{16.10}$$

We can then express the redshift-dependent cosmic-ray injection rate in terms of the local rate, and the ratio of the redshift-dependent comoving rate to that in the local universe, $f(z)$:

$$E_{\rm CR}^2 \frac{d\dot{N}_{\rm CR}}{dE_{\rm CR}}(z) = E_{\rm CR}^2 \frac{d\dot{N}_{\rm CR}}{dE_{\rm CR}}\bigg|_{z=0} \times f(z). \tag{16.11}$$

We generally expect $f(z)$ to be a function of order unity (and defined to be unity at $z = 0$), reflecting the evolution of the comoving number density of ultra-high energy cosmic-ray sources over cosmological history.

From the injection rate of cosmic rays, we can calculate the flux of neutrinos that arises from their interactions:

$$E_\nu^2 \frac{dN_\nu}{dE_\nu} = \int E_{\rm CR}^2 \frac{d\dot{N}_{\rm CR}}{dE_{\rm CR}}(z)\, \epsilon\, f_{\pi^\pm} f_\nu \, \frac{c}{4\pi} \, \frac{dt}{dz} \, \frac{dz}{(1+z)}, \tag{16.12}$$

where ϵ is the fraction of the energy in these cosmic rays that goes into the production of pions, $f_{\pi^\pm}$ is the fraction of the pions that are charged, and f_ν is the fraction of the energy in charged pions that goes into neutrinos. The conversion factor between time and redshift can be written in terms of the Hubble rate, $dt/dz = -1/[H(z)(1+z)]$. Finally, the factor of $(1+z)$ in the denominator accounts for the effects of cosmological redshift.

Note that the neutrinos that are produced in these interactions carry only a fraction of the energy of their parent cosmic ray. For protons, this ratio is typically around $E_\nu \sim 0.05 \times E_p$. This gap in energy is even larger for cosmic-ray nuclei, which first undergo photodisintegration before the resulting nucleons can go on to produce pions and neutrinos. For this reason, this calculation reflects the neutrino spectrum that we expect to be present at somewhat lower energies, roughly around $E_\nu \sim 10^7$–10^{10} GeV.

Using the information described above, we can rewrite Eq. 16.12 as follows:

$$\begin{aligned} E_\nu^2 \frac{dN_\nu}{dE_\nu} &= E_{\rm CR}^2 \frac{d\dot{N}_{\rm CR}}{dE_{\rm CR}}\bigg|_{z=0} \frac{\epsilon\, f_{\pi^\pm} f_\nu c}{4\pi} \int_0^{z_{\rm max}} \frac{f(z)}{H(z)(1+z)^2}\, dz \\ &= E_{\rm CR}^2 \frac{d\dot{N}_{\rm CR}}{dE_{\rm CR}}\bigg|_{z=0} \frac{\epsilon\, f_{\pi^\pm} f_\nu c}{H_0 4\pi} \int_0^{z_{\rm max}} \frac{f(z)}{[\Omega_M (1+z)^3 + \Omega_\Lambda]^{1/2}(1+z)^2}\, dz. \end{aligned} \tag{16.13}$$

For any reasonable distribution of cosmic-ray sources, the integral in this expression will be of order unity. For sources that are distributed with a constant comoving number density, for example, this integral yields approximately 0.53. Traditionally, the impact of the source distribution on this result has been quantified in terms of the parameter ξ, which is related to the integral in the above expression as follows:

$$\xi \equiv \frac{\int_0^{z_{\rm max}} f(z)\, H(z)^{-1}(1+z)^{-2}\, dz}{\int_0^\infty H(z)^{-1}(1+z)^{-1}\, dz}. \tag{16.14}$$

Most of the proposed sources of the high-energy cosmic rays follow distributions that correspond to $\xi \sim 1 - 6$.

Using the local cosmic-ray injection rate given in Eq. 16.10, this line of reasoning yields the following flux of high-energy neutrinos:

$$\begin{aligned} E_\nu^2 \frac{dN_\nu}{dE_\nu} &= E_{\rm CR}^2 \frac{d\dot{N}_{\rm CR}}{dE_{\rm CR}}\bigg|_{z=0} \frac{\epsilon\, f_{\pi^\pm} f_\nu c}{4\pi}\, \xi \int_0^\infty \frac{dz}{H(z)(1+z)} \\ &\approx 2.5 \times 10^{-8}\ {\rm GeV/cm^2/s/sr} \times \left(\frac{\epsilon}{1}\right)\left(\frac{f_{\pi^\pm}}{0.5}\right)\left(\frac{f_\nu}{0.75}\right)\left(\frac{\xi}{1}\right). \end{aligned} \tag{16.15}$$

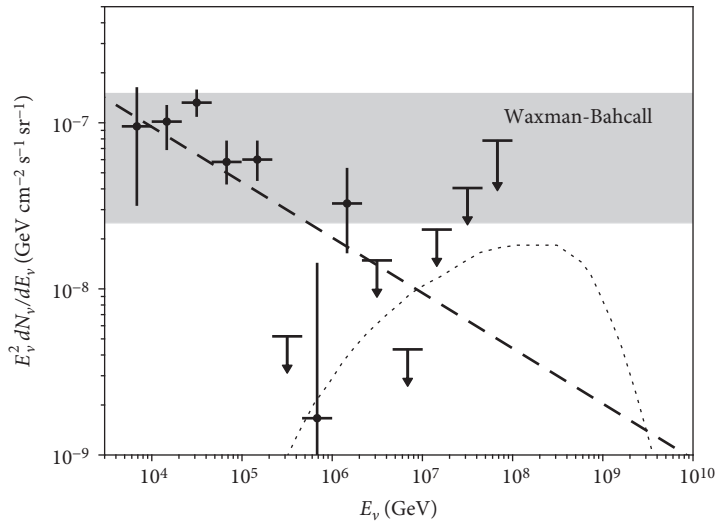


Figure 16.3 The diffuse high-energy neutrino flux measured by IceCube [299, 300, 301], compared to the Waxman-Bahcall bound (for $\xi = 1{-}6$) [297, 298], and to the predictions of a non-blazar AGN model (dashed) and the blazar model of Ref. [302] (dotted).

This famous result [297, 298] provides us with a model-independent estimate for the diffuse flux of high-energy neutrinos from extragalactic sources of cosmic rays. For the specific case of $\epsilon = 1$, this result is known as the Waxman-Bahcall bound, which is the upper limit on the diffuse flux of high-energy neutrinos from optically thin sources of cosmic rays.

The single greatest uncertainty in this calculation comes from the fraction of the energy in cosmic rays that goes into the production of pions, ϵ. As we will see later in this section, many potential sources of high-energy cosmic rays lead to estimates of $\epsilon \sim 0.01$–1, providing us with strong motivation to build and operate telescopes that are sensitive to neutrino fluxes near the Waxman-Bahcall bound. One should also keep in mind that ϵ could vary with energy, especially in the case of neutrinos that are produced through proton-photon interactions.

The IceCube experiment has detected a diffuse flux of neutrinos with a spectrum that extends up to at least several PeV [299, 300, 301] (see Figure 16.3). While the origin of these neutrinos is not yet clear, their angular distribution shows no significant departures from isotropy, indicating that they are produced by a large number of extragalactic sources. Furthermore, the neutrino flux that is observed at PeV-scale energies is roughly equal to that predicted for $\epsilon \sim 0.1 - 1$ (for a reasonable range of source evolution scenarios, $\xi \sim 1 - 6$). This suggests that a significant fraction of the energy in $\sim 10^8$ GeV cosmic rays is indeed converted into pions, resulting in a flux of neutrinos that is not far below the Waxman-Bahcall bound.

16.4.2 Active galactic nuclei

Nearly all massive galaxies harbor a supermassive black hole at their center. While most of these objects are more or less quiescent, a small fraction ($\sim 1 - 10\%$ [303]) are extremely bright, being powered by the ongoing accretion of matter onto their central black hole. Such objects are known as active galactic nuclei (AGN), and are the most luminous persistent (as opposed to transient) sources of electromagnetic radiation in our universe.

AGN constitute the majority of all detected gamma-ray sources [304] and, as discussed in Sec. 15.4, are among the most promising candidates for the sources of the highest energy cosmic rays [305].

From an observational perspective, there are many different kinds of AGN. This is the consequence of physical distinctions, but also of geometrical considerations. In particular, many (although not all) AGN feature highly relativistic and highly luminous jets, causing their appearance to depend strongly on the orientation of those jets relative to the observer [306]. Blazars, for example, are AGN with a jet that is pointed within an angle of roughly $\theta \sim 1/\gamma$ of our direction, where γ is the Lorentz factor of the gas within the jet (also known as the bulk Lorentz factor of the jet) [307]. Blazars are the brightest class of AGN, and feature highly variable emission in the optical, radio, X-ray, and gamma-ray bands. Blazars are further classified as either BL Lacertae objects (BL Lacs) or flat-spectrum radio quasars (FSRQs), depending on the detailed characteristics of their observed emission.

The properties of non-blazar AGN are also highly varied, depending on factors such as whether they have jets, the alignment of those jets, the rate of accretion, the type of host galaxy, and the amount of gas that is present in the galaxy's center. Depending on these and other features, AGN are often divided into different subclassifications, including radio galaxies, Seyfert galaxies, and quasars, among many others.

In the case of blazars, high-energy neutrinos could be produced through the scattering of high-energy protons with photons in the jet. Key to this picture are the large Lorentz factors of the radiation and other material that is traveling down the jet in the direction of the observer. We can use the fact that $E_\gamma \gtrsim 10$ TeV gamma rays are frequently observed from blazars to place a lower limit on this Lorentz factor. In particular, the threshold for a gamma ray to undergo pair production in the jet is approximately given by $E_\gamma \epsilon_\gamma \gtrsim (\gamma m_e)^2$, where γ is the Lorentz factor and ϵ_γ is the energy of a target photon (see Sec. 5.3). Since the observed gamma rays must be below this threshold, we can infer the following constraint:

$$\begin{aligned} \gamma &\gtrsim \left(\frac{E_\gamma \epsilon_\gamma}{m_e^2}\right)^{1/2} \\ &\sim 30 \times \left(\frac{E_\gamma}{20\,\text{TeV}}\right)^{1/2} \left(\frac{\epsilon_\gamma}{10\,\text{eV}}\right)^{1/2}. \end{aligned} \tag{16.16}$$

The gamma-ray emission from blazars is often highly variable, featuring flares that persist for minutes or hours. Consider such a flare, with a total luminosity L_γ and a duration Δt. Treating the boosted material that constitutes this flare as roughly spherical, we can write the energy density in the moving frame as follows:

$$\begin{aligned} \rho'_\gamma &= \frac{L'_\gamma \Delta t}{\frac{4}{3}\pi r'^3} \\ &= \frac{(L_\gamma/\gamma)\Delta t}{\frac{4}{3}\pi(\gamma c \Delta t)^3} \\ &= \frac{3 L_\gamma}{4\pi \gamma^4 c^3 \Delta t^2}. \end{aligned} \tag{16.17}$$

The primed quantities in this expression correspond to values in the moving frame of reference, while the unprimed ones are as they appear in the observer's frame. For example,

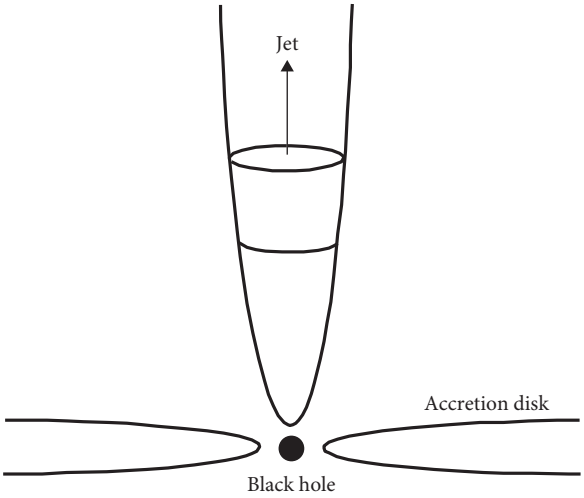


Figure 16.4 A schematic diagram of an AGN jet.

L_γ is the observed luminosity of the blazar, while $L'_\gamma = L_\gamma/\gamma$ is the luminosity in the moving frame. Similarly, r is the radius of the region in the moving frame and $r' = \gamma c \Delta t$ is the corresponding length in the observer's frame.

We are now in a position to estimate how many of the cosmic rays in the jet will produce pions, and thus gamma rays and neutrinos. The fraction of the energy in these cosmic rays that goes into the production of pions is given by the following:

$$\begin{aligned}
\epsilon &= \frac{r'}{\lambda_{p\gamma}} \langle x_{p\to\pi} \rangle \qquad (16.18)\\
&= (\gamma c \Delta t)\, n'_\gamma \sigma_{p\gamma} \langle x_{p\to\pi} \rangle \\
&= (\gamma c \Delta t)\, \frac{\rho'_\gamma}{\epsilon'_\gamma} \sigma_{p\gamma} \langle x_{p\to\pi} \rangle \\
&= (\gamma c \Delta t)\, \frac{3L_\gamma}{4\pi \gamma^4 c^3 \Delta t^2}\, \frac{1}{\epsilon_\gamma/\gamma}\, \sigma_{p\gamma} \langle x_{p\to\pi} \rangle \\
&= \frac{3L_\gamma}{4\pi \gamma^2 c^2 \Delta t\, \epsilon_\gamma}\, \sigma_{p\gamma} \langle x_{p\to\pi} \rangle \\
&\sim 0.6 \times \left(\frac{L_\gamma}{10^{38}\,\mathrm{GeV}}\right) \left(\frac{10\,\mathrm{eV}}{\epsilon_\gamma}\right) \left(\frac{30}{\gamma}\right)^2 \left(\frac{10^3\,\mathrm{s}}{\Delta t}\right),
\end{aligned}$$

where $\lambda_{p\gamma} = 1/(n'_\gamma \sigma_{p\gamma})$ is the mean free path of a proton in the moving frame and $\langle x_{p\to\pi} \rangle \sim 0.2$ is the average fraction of a colliding proton's energy that is transferred into pions. In the last step, I have taken $\sigma_{p\gamma} \sim 10^{-28}\,\mathrm{cm}^2$, which is appropriate for energies near the Δ-resonance (see Figure 5.3). Note that in a more precise calculation, we would integrate over the energies of the target radiation and account for the energy dependence of the inelastic cross section, $\sigma_{p\gamma}$.

This calculation demonstrates that a significant fraction of the high-energy protons that are present in a blazar jet should be expected to produce pions, at least for this choice of parameters, leading to a flux of high-energy neutrinos that is not far below the Waxman-Bahcall bound. This will only be the case, however, for those protons that have enough energy to produce pions. For this to occur through the Δ-resonance, the total energy in the center-of-momentum frame of the collision must satisfy

$$E_{\rm CM} = (4E'_p\epsilon'_\gamma)^{1/2} \gtrsim (m_\Delta^2 - m_p^2)^{1/2}, \tag{16.19}$$

which, in the frame of the observer, corresponds to

$$\begin{aligned} E_p &\gtrsim \frac{(m_\Delta^2 - m_p^2)\gamma^2}{4\epsilon_\gamma} \\ &\sim 1.4 \times 10^{10}\,{\rm GeV} \times \left(\frac{\gamma}{30}\right)^2 \left(\frac{10\,{\rm eV}}{\epsilon_\gamma}\right). \end{aligned} \tag{16.20}$$

Protons above this threshold produce neutrinos with energies in excess of the following:

$$\begin{aligned} E_\nu &\sim \frac{1}{4}\langle x_{p\to\pi}\rangle E_p \\ &\gtrsim 7 \times 10^8\,{\rm GeV} \times \left(\frac{\gamma}{30}\right)^2 \left(\frac{10\,{\rm eV}}{\epsilon_\gamma}\right), \end{aligned} \tag{16.21}$$

where the factor of 1/4 reflects the fact that each charged pion decays to four particles, and thus each neutrino carries approximately a quarter of the pion's energy.

In summary, blazar jets seem to be environments in which ultra-high energy protons could quite efficiently produce very high-energy neutrinos. The total neutrino flux that is predicted to arise from this class of sources depends on many details of the model, including the spectrum and density of the target radiation, the characteristics of the flares, and the extent to which these objects are, in fact, the sources of the high and ultra-high energy cosmic rays [308, 302, 309, 310]. In models in which blazars are the main sources of the highest energy cosmic rays, the neutrino spectrum from these objects is typically predicted to peak at PeV-EeV energies, with fluxes that are roughly equal to the Waxman-Bahcall bound [309, 308, 302].

Empirically speaking, the lack of any observed correlation between the directions of high-energy neutrinos and known blazars (with the possible exception of the blazar TXS 0506+056 [311, 312]) has been used to place upper limits on the fraction of the diffuse neutrino flux observed by IceCube that could arise from this class of objects. Such studies find that no more than $\sim$15% of these neutrinos can originate from blazars [313, 314, 315, 316]. At higher energies, however, it remains possible that blazars could dominate the astrophysical neutrino spectrum.

Blazars are not the only class of AGN that could potentially produce high-energy neutrinos. It is plausible that AGN jets accelerate high and ultra-high energy cosmic rays, but that many or most of these particles escape the jet and then diffuse through the volume of their parent galaxy (or galaxy cluster) for some length of time. During that time, they would scatter with gas, producing pions and, therefore, high-energy neutrinos. In such a scenario, a large fraction of AGN, not limited to blazars, could potentially contribute to the observed spectrum of astrophysical neutrinos.

Unlike the proton-photon scattering that is most important within the jet of an AGN, proton-proton scattering typically leads to a power-law spectrum of neutrinos, with a spectral index that is closely related to that of the cosmic-ray protons. The fraction of the energy in these cosmic rays that goes into the production of pions in this case is given by $\epsilon \sim \sigma_{pp}\, c\, t_{\rm esc}\, n_{\rm gas}$, where $t_{\rm esc}$ is the timescale on which the cosmic rays escape from the region, and $n_{\rm gas}$ is the number density of gas. Treating cosmic-ray transport as diffusive, these particles travel a distance, $L_{\rm dif}$, to the edge of the diffusive region in time given by $t_{\rm esc} \sim (L_{\rm dif}/2)^2/D$, where $D = D_0 E_p^{\alpha}$ is the diffusion coefficient (see Sec. 15.2). Using this information, and adopting diffusion parameters similar to those measured in the Milky Way, $D_0 \approx 4 \times 10^{28}\,{\rm cm}^2/{\rm s}$ and $\alpha \approx 1/3$, the fraction of the energy in cosmic rays that goes into pion production is given by

$$\begin{aligned} \epsilon &\sim \sigma_{pp}\, c\, t_{\rm esc}\, n_{\rm gas} \\ &\sim 0.06 \times \left(\frac{L_{\rm dif}}{10\,{\rm kpc}}\right)^2 \left(\frac{1\,{\rm PeV}}{E_p}\right)^{1/3} \left(\frac{n_{\rm gas}}{0.2\,{\rm cm}^{-3}}\right). \end{aligned} \tag{16.22}$$

Recall from Sec. 16.2.2 that proton-proton scattering is expected to lead to a spectrum of gamma rays and neutrinos that is somewhat softer than that of the parent protons. In particular, high-energy protons with an index of $\Gamma_p \approx 2$ will produce a spectrum of neutrinos with $\Gamma_\nu \approx 4\Gamma_p/3 - 2/3 + 1/3 \approx 2.3$, where the final 1/3 is the result of diffusion. Comparing this to the diffuse neutrino spectrum reported by IceCube, this class of sources seems to provide at least a plausible explanation for the observed characteristics of this signal.

In Figure 16.3, I show the neutrino flux that is predicted from proton-proton collisions in and around non-blazar AGN, for a spectral index of $\Gamma_\nu = 2.3$, normalized to $\epsilon = 0.06$ at $E_\nu = 1\,{\rm PeV}$, and with a redshift distribution corresponding to $\xi = 5$. Also shown is an example of the neutrino spectrum that is predicted from a particular blazar model [302], normalized to $\epsilon = 0.14$ at its peak energy, and adopting the same redshift distribution.

16.4.3 Gamma-ray bursts

Gamma-ray bursts (GRB) are extreme explosions that typically feature luminosities in the range of $\sim 10^{53} - 10^{57}\,{\rm GeV/s}$ over a period of roughly $\sim 1 - 100$ seconds, making them the most energetic and luminous events in the universe [317, 318, 319]. GRB are typically subdivided into long-duration ($\gtrsim 2\,{\rm sec}$) bursts, which are thought to originate from the collapse of a massive star [320], and short-duration ($\lesssim 2\,{\rm sec}$) bursts, which are generally associated with the merger of two neutron stars, or a neutron star and a black hole [321, 322, 323].

In a GRB, plasma is ejected from the central engine in a series of successive shells. As the outer shells slow down, collisions between shells take place, leading to the formation of internal shock fronts that could potentially be capable of accelerating cosmic rays up to the highest observed energies [324, 325]. As these shells expand, the plasma becomes increasingly optically thin, leading to the production of a bright burst of gamma rays. An "afterglow" of longer wavelength emission can also be observed from GRB, persisting for as long as months, and arising from the collisions of the shells with the surrounding interstellar medium.

In the initial configuration of a GRB, there is a radiation-dominated bath of electrons and photons, along with a few nucleons, not unlike the early universe. The radiation pressure that is present in this state causes this material to expand outward in highly relativistic

shells. Protons can scatter with the gamma rays in these shells to produce pions through the Δ-resonance, but to do so they must satisfy (see Eq. 16.20)

$$E_p \gtrsim \frac{(m_\Delta^2 - m_p^2)\gamma^2}{4\epsilon_\gamma} \tag{16.23}$$
$$\sim 1.4 \times 10^7 \text{ GeV} \times \left(\frac{\gamma}{300}\right)^2 \left(\frac{1 \text{ MeV}}{\epsilon_\gamma}\right),$$

leading to neutrinos with energies above

$$E_\nu \sim \frac{1}{4}\langle x_{p\to\pi}\rangle E_p \tag{16.24}$$
$$\gtrsim 7 \times 10^5 \text{ GeV} \times \left(\frac{\gamma}{300}\right)^2 \left(\frac{1 \text{ MeV}}{\epsilon_\gamma}\right).$$

As we did when considering proton-photon collisions within blazar jets, we can now estimate the fraction of the energy in the cosmic rays accelerated by GRB that goes into the production of pions. The geometry in this case, however, is somewhat different, featuring a shock at radius r of width Δr. In the rest frame of the GRB fireball, these quantities are given by $r' = \gamma^2 c\Delta t$ and $\Delta r' = \gamma c \Delta t$, leading to the following efficiency for pion production:

$$\epsilon = \Delta r'\, n'_\gamma \sigma_{p\gamma}\, \langle x_{p\to\pi}\rangle \tag{16.25}$$
$$= \Delta r'\, \frac{\rho'_\gamma}{\epsilon'_\gamma}\, \sigma_{p\gamma}\, \langle x_{p\to\pi}\rangle$$
$$= \Delta r'\, \frac{L_\gamma \Delta t/\gamma}{4\pi r'^2 \Delta r'} \frac{1}{(\epsilon_\gamma/\gamma)}\, \sigma_{p\gamma}\, \langle x_{p\to\pi}\rangle$$
$$= \frac{L_\gamma}{4\pi\gamma^4 c^2 \Delta t \epsilon_\gamma}\, \sigma_{p\gamma}\, \langle x_{p\to\pi}\rangle$$
$$\approx 0.2 \times \left(\frac{L_\gamma}{10^{55} \text{ GeV/s}}\right) \left(\frac{1 \text{ MeV}}{\epsilon_\gamma}\right) \left(\frac{300}{\gamma}\right)^4 \left(\frac{10^{-2} \text{ s}}{\Delta t}\right).$$

Note that variations in GRB light curves are observed on very short timescales, indicating that $\Delta t \sim \mathcal{O}(1\text{–}100 \text{ ms})$. Also notice that this result depends very strongly on the value of the Lorentz factor, γ.

For reasonable choices of the relevant parameters, the neutrino flux from GRB could approximately saturate the Waxman-Bahcall bound, especially at energies in the range of $E_\nu \sim 10^5\text{–}10^8$ GeV [326, 327, 328, 329]. That being said, the high-energy neutrino events observed by IceCube have not been found to correlate in either time or direction with known GRB [330], indicating that most of these particles do not originate from this class of objects.

16.4.4 Cosmogenic neutrinos

Predictions for the fluxes of high-energy neutrinos from AGN and GRB depend critically on whether these objects accelerate a significant fraction of the high and ultra-high energy cosmic rays, and on the fraction of those cosmic rays that undergo pion production before

escaping from their sources. As a consequence, these predictions feature large uncertainties. On the other hand, we have measured the cosmic-ray spectrum, and can be confident that these particles scatter on known radiation fields, such as the CMB and the cosmic infrared background, leading to the production of pions and neutrinos. This mechanism should, therefore, be guaranteed to generate a flux of high-energy "cosmogenic" neutrinos [331, 332, 333].

The spectrum of the cosmogenic neutrinos depends on the detailed spectrum and chemical composition of the ultra-high energy cosmic rays, as well as on the redshift distribution of their sources. Beginning with the case of ultra-high energy protons, such particles can scatter with CMB photons through the Δ-resonance so long as their energies exceed the following threshold:

$$E_p > \frac{(m_\Delta^2 - m_p^2)}{4\epsilon_{\rm CMB}} \sim 2.5 \times 10^{11}\,{\rm GeV} \times \left(\frac{6.3 \times 10^{-4}\,{\rm eV}}{\epsilon_{\rm CMB}}\right). \tag{16.26}$$

Integrating over the energy distribution of CMB photons, protons with energies above $E_p \gtrsim 10^{10}$ GeV undergo pion production efficiently over cosmological distances, leading to a spectrum of neutrinos that peaks at around the following energy:

$$E_\nu \sim \frac{\langle x_{p\to\pi}\rangle}{4} E_p \sim 5 \times 10^8\,{\rm GeV}. \tag{16.27}$$

Because pion production is so efficient at energies above this threshold, the cosmogenic flux is predicted to roughly saturate the Waxman-Bahcall bound in the all-proton case [334, 335]. Empirically speaking, however, it is clear that most of the highest energy cosmic rays are not protons, but instead consist of a mixture of nuclei. In order for cosmic-ray nuclei to generate neutrinos, they must first undergo photo-disintegration, creating protons which can then go on to produce pions and neutrinos.

The photo-disintegration of ultra-high energy nuclei occurs primarily through interactions with the cosmic infrared background, leading to scattering rates that are quite efficient over cosmological distance scales [336, 281, 337]. The energies of the resulting nucleons, however, are smaller than those of their parent nuclei by a factor of the nucleus' atomic mass. For heavy nuclear species, this can often lead to nucleons that have too little energy to produce pions. Ultra-high energy cosmic-ray models that provide a good fit to the current data generally predict cosmogenic neutrino fluxes that are significantly suppressed relative to the all-proton case, falling well below the Waxman-Bahcall bound, and peaking at somewhat lower energies [338, 339, 340, 341].

16.4.5 Supernova remnants and pulsars

Supernovae are extremely powerful and luminous events, triggered by the collapse of a massive star, or by the reignition of fusion in a degenerate star, such as a white dwarf. These events often result in the formation of a neutron star or a black hole. At their peak luminosity, supernovae can produce more energy in light than their entire host galaxy. Furthermore, these explosions can expel up to several solar masses of material at speeds on the order of $v \sim 10^4$ km/s.

The cores of supernovae typically reach temperatures of about $T \sim 30$ MeV, allowing them to produce large numbers of thermal neutrinos. Unlike photons, which are efficiently trapped within such objects, neutrinos can much more readily escape. As a result, supernovae typically emit on the order of $\sim 10^{58}$ neutrinos and antineutrinos, which carry away roughly 99% of the system's total gravitational energy [342, 343]. It is estimated that supernovae occur in the Milky Way at a rate of approximately three per century. The most recent example of such an event was SN 1987A, which occurred in the Large Magellanic Cloud (a satellite galaxy of the Milky Way). The Kamiokande II, IMB, and Baksan 5 detectors collectively observed 25 neutrinos associated with SN 1987A, consistent with expectations.

At energies up to the knee, the cosmic-ray spectrum is thought to originate largely from young supernova remnants. Such objects feature an expanding supersonic shock wave that collides with and sweeps up the material in the surrounding interstellar medium. This class of objects constitutes one of the main sources of TeV-scale gamma-ray emission, with evidence of contributions from both inverse Compton scattering and pion production [344, 345, 346, 347]. Supernova remnants are also expected to be potentially detectable sources of TeV-scale neutrino emission, providing a clear test of their hadronic cosmic-ray acceleration [332, 348, 349].

When a star collapses to form a neutron star, its radius decreases by a factor of $r_i/r_f \sim 10^5$. From the conservation of angular momentum, this causes the rate at which the star is rotating to increase by a factor of $(r_i/r_f)^2 \sim 10^{10}$, leading to periods on the order of milliseconds. This sudden collapse also increases the star's dipolar magnetic field by a factor proportional to $(r_i/r_f)^2$, resulting in typical values of $B \sim 10^{11}$–10^{13} G. Newly formed neutron stars are thus both highly magnetized and very rapidly spinning.

Being rotating magnetic dipoles, young neutron stars emit electromagnetic radiation. This radiation takes the form of radio and gamma-ray beams that periodically sweep out portions of the surrounding solid angle. To external observers, these rotating beams take on the appearance of a sequence of regular flashes, earning this class of objects the name of pulsars.

The neutron star at the heart of a pulsar gradually loses its rotational kinetic energy, at a rate that can be expressed as follows:

$$\begin{aligned} \frac{dE_{\rm rot}}{dt} &= \frac{d}{dt}\left(\frac{1}{2} I\omega^2\right) \\ &= -\frac{4\pi^2 I}{P^3}\frac{dP}{dt}, \end{aligned} \tag{16.28}$$

where I, ω, and P are the pulsar's moment of inertia, rotational frequency, and period.

According to the conservation of energy, this loss of rotational kinetic energy must be equal to the total power that is emitted from a pulsar, including that in the form of radio and gamma-ray emission, as well as electron-positron pairs. If this emission is dominated by magnetic dipole radiation, its power can be expressed as

$$P_{\rm rad} = \frac{2(BR^3 \sin\alpha)^2}{3c^3}\left(\frac{2\pi}{P}\right)^4, \tag{16.29}$$

where R is the radius of the neutron star, α is the angle between the magnetic and rotational axes of the pulsar, and $BR^3 \sin\alpha$ is the perpendicular component of the magnetic dipole.

Combining these expressions, we can express dP/dt as

$$\frac{dP}{dt} = -\frac{8\pi^2 (BR^3 \sin\alpha)^2}{3c^3 IP}, \tag{16.30}$$

and $dE_{\rm rot}/dt$ as

$$\begin{aligned}\frac{dE_{\rm rot}}{dt} &= \frac{32\pi^4 B^2 R^6 \sin^2\alpha}{3c^3 P^4} \\ &\approx 4\times 10^{37}\ {\rm GeV/s} \times \left(\frac{B}{10^{12}\,{\rm G}}\right)^2 \left(\frac{R}{15\,{\rm km}}\right)^6 \left(\frac{\sin\alpha}{0.5}\right)^2 \left(\frac{0.2\,{\rm s}}{P}\right)^4.\end{aligned} \tag{16.31}$$

In deriving this result, we assumed that the emission from a pulsar is due entirely to magnetic dipole radiation. If this is not the case, the pulsar's rotational kinetic energy will decline more rapidly. With this in mind, one often introduces a quantity known as the pulsar braking index, n, which can be defined as $dP/dt \propto -P^{-n+2}$. While many individual pulsars have been observed to have braking indices close to $n = 3$ (corresponding to the case of pure magnetic dipole breaking), others exhibit evolution as rapid as $n = 1.4$ [350].

From this relationship between dP/dt and P, we can deduce the functional form for the time evolution of the pulsar's period:

$$\begin{aligned}&\frac{dP}{dt} = -CP^{-n+2} \\ &\longrightarrow \int_{P_0}^{P(t)} P^{n-2}\, dP = -C\int_0^t dt' \\ &\longrightarrow P(t) = P_0\left(\frac{C\,(1-n)\,t}{P_0^{n-1}} + 1\right)^{1/(n-1)},\end{aligned} \tag{16.32}$$

where $C = -(dP/dt)/P^{-n+2}$ is a constant of proportionality and P_0 is the initial period of the pulsar. We can also express this in terms of the spindown timescale, τ:

$$P(t) = P_0\left(\frac{t}{\tau} + 1\right)^{1/(n-1)}, \tag{16.33}$$

where

$$\tau = \frac{P}{(dP/dt)(n-1)}\left(\frac{P_0}{P}\right)^{n-1}. \tag{16.34}$$

It is sometimes convenient to define the "characteristic age" of a pulsar as $t_c \equiv P/2(dP/dt)$, allowing us to relate

$$\tau = \frac{2t_c}{n-1}\left(\frac{P_0}{P}\right)^{n-1}. \tag{16.35}$$

Note that the true age of a pulsar is related to its characteristic age by $t_{\rm age} = t_c + \tau$.

Putting this together, we can now express the time evolution of $dE_{\rm rot}/dt$ as follows:

$$\begin{aligned}\frac{dE_{\rm rot}}{dt} &= -\frac{4\pi^2 I}{P^3}\frac{dP}{dt} \\ &= -\frac{4\pi^2 I}{P_0^2(n-1)\tau}\left(\frac{t}{\tau}+1\right)^{-(n+1)/(n-1)}.\end{aligned} \tag{16.36}$$

The spindown timescale, τ, corresponds to the time at which a pulsar's period ceases to be approximately constant, and instead begins to evolve as $P \propto t^{1/(n-1)}$. The timing of this transition depends on the pulsar's magnetic field and initial period. In the case of magnetic dipole braking ($n = 3$), the spindown timescale is given by

$$\begin{aligned}\tau &= \frac{3c^3 I P_0^2}{4\pi^2 B^2 R^6} \\ &\approx 1.3\times 10^4\,{\rm yr}\times\left(\frac{10^{12}\,{\rm G}}{B}\right)^2\left(\frac{M}{1.4\,M_\odot}\right)\left(\frac{15\,{\rm km}}{R}\right)^4\left(\frac{P_0}{0.03\,{\rm s}}\right)^2.\end{aligned} \tag{16.37}$$

Pulsar beams include significant emission at both radio and gamma-ray wavelengths. The gamma-ray beams from these objects feature a rather hard spectral index with an exponential cutoff above a few GeV, $dN_\gamma/dE_\gamma \propto E^{-1.6}\times\exp(-E_\gamma/3.8\,{\rm GeV})$ [351]. In addition, pulsars also transfer a significant fraction of their spindown power into the production of very high-energy electron-positron pairs. These particles slowly diffuse throughout the region surrounding the pulsar, losing energy and producing very high-energy gamma-ray emission through inverse Compton scattering, giving rise to what are known as "TeV halos" [285, 352].

Pulsars that are created directly through the process of neutron star formation typically lose their rotational kinetic energy fairly quickly, with $\tau \sim \mathcal{O}(10^4{\rm yr})$. In addition to these young pulsars, long dormant neutron stars can in some cases gain angular momentum from a companion star, giving rise to pulsars that have millisecond-scale periods but relatively weak magnetic fields, $B \sim 10^7$–10^8 G. In contrast to young pulsars, these "recycled" or "millisecond" pulsars evolve much more slowly, retaining their rotational kinetic energy for hundreds of millions or even billions of years.

Additional reading

I've long used Longair's *High-Energy Astrophysics* as a primary reference in the areas of gamma-ray and neutrino astronomy. For reviews of high-energy neutrino astronomy, I often turn to those by Becker [353], and by Halzen and Hooper [354].

Problems

16.1 From the fact that proton-proton collisions produce approximately $N_\pi \approx 3\times[(E_p - E_{\rm th})/{\rm GeV}]^{1/4}$ pions, it follows that the average energy of these pions will be given by $\langle E_\pi\rangle \propto E_p^{3/4}$ (in the $E_p \gg m_p$ limit). Use this fact to show that the collisions of a power-law spectrum of cosmic-ray protons with gas will give rise

to a power-law spectrum of neutrinos and gamma rays with a spectral index of $\Gamma_{\nu,\gamma} \approx (4/3)\Gamma_p - (2/3)$, where Γ_p is the index of the proton spectrum, $dN_p/dE_p \propto E_p^{-\Gamma_p}$. Assume that every pion is created with the average predicted energy.

16.2 Consider a high-energy photon traveling a distance of 1 kpc through the local neighborhood of our galaxy. In addition to the CMB, this photon could scatter with the optical light from stars, which has an energy density of $\rho_{\rm star} \sim 0.3$ eV/cm^3. As a function of energy, estimate the probability that this photon will undergo pair production.

16.3 Consider a 1 EeV photon that is emitted from a source located at a distance of 1 Gpc. Estimate how low in energy the resulting electromagnetic cascade will develop before reaching us. For simplicity, assume that each interaction occurs at its mean free path, and that those interactions produce particles with the average expected energy.

16.4 Calculate the oscillation probabilities as given in Eq. 16.8, but taking $\sin^2\theta_{13} = 0.022$ (instead of zero). How much does this change the results?

16.5 Consider a relatively exotic astrophysical source that produces high-energy antineutrinos mostly through the decays of highly boosted neutrons, $n \to p + e^- + \bar{\nu}_e$. After traveling a long distance, what will the flavor ratio of these antineutrinos be?

16.6 In our discussion of high-energy neutrinos, we have focused on their production through the decays of charged pions. How important might kaon decays be to the process of high-energy neutrino production in astrophysical environments?

16.7 Can you think of any ways that nature could produce a flux of high-energy astrophysical neutrinos that is greater than that of the Waxman-Bahcall bound?

16.8 Consider a class of cosmic-ray sources whose number density evolves proportionally to the star formation rate, $f(z) = (1 + 7.6z)/[1 + (z/3.3)^{5/3}]$ [355]. For this case, calculate the source evolution parameter, ξ.

16.9 The Kamiokande II, IMB, and Baksan 5 detectors collectively observed 25 neutrinos from supernova 1987A. Based on the size of these detectors, estimate how large a similar detector would have to be in order to detect one neutrino from a type II supernova located in the Andromeda galaxy.

16.10 Compare the spindown power, $dE_{\rm rot}/dt$, of a typical young pulsar with $P = 2$ s and $B = 10^{12}$G to that of a millisecond pulsar with $P = 0.01$ s and $B = 10^8$ G.

17

Particle Dark Matter Candidates

In Chapter 9, I discussed the subject of dark matter while remaining largely agnostic about the specific particle nature of this substance. In this chapter, I instead focus on a few examples of particle physics frameworks from which particularly well-motivated candidates for dark matter can arise. These dark matter candidates include neutralinos in supersymmetric models, sterile neutrinos, and axions, as well as particles that can appear within the context of hidden sector models. For each of these dark matter candidates, I summarize their properties, review the relevant constraints, and discuss how they may have been produced in the early universe.

17.1 Supersymmetric dark matter

Of all of the extensions of the Standard Model to have been proposed over the years, perhaps none have received as much attention as supersymmetry [356, 357, 358, 359]. The basic idea of supersymmetry is that for each fermion, there exists a boson—its superpartner—with identical gauge charges and other couplings. Thus, if supersymmetry is realized in nature, fermions, such as electrons, neutrinos, and quarks, must each be accompanied by a pair of scalar superpartners: selectrons, sneutrinos, and squarks. In turn, the photon and gluon must be partnered to the fermionic photino and gluino. Since no supersymmetric particles have been observed at any collider, or in any other experiment, supersymmetry must be a broken symmetry, enabling the superpartners to be significantly heavier than their Standard Model counterparts.

In complete generality, the symmetries that could be manifest within a physical theory can be divided into two categories. First, there are those symmetries that leave a physical theory unchanged under operations in spacetime (translations, rotations, etc.). Through Noether's theorem, these symmetries are associated with the conservation of quantities such as momentum, angular momentum, and energy. Second, there are internal symmetries, such as gauge symmetries, which are central to modern quantum field theory, and which are responsible for the conservation of electric charge and color. Supersymmetry is unique in that, under reasonable assumptions (locality, etc.), it is the *only* possible example of a

symmetry that is both an internal symmetry, and a symmetry of spacetime.[1] While this does not guarantee that supersymmetry is manifest in nature, its uniqueness does make it stand out as a particularly interesting topic of study.

Phenomenologically speaking, the inclusion of supersymmetry within the laws of physics has the potential to be able to solve a number of long-standing problems in both particle physics and cosmology. These motivations include the following:

- **Stabilization of the Higgs mass**: In the Standard Model, radiative corrections should be expected to dramatically increase the mass of the Higgs boson to well above its measured value. The existence of superpartners with masses not too far above the electroweak scale could lead to the cancellation of the largest of these corrections, explaining how the Higgs could be so relatively light. I'll discuss this issue—known as the hierarchy problem—in more detail in Sec. 17.1.1.
- **Gauge coupling unification**: In grand unified theories (GUTs), the $SU(3)_C \times SU(2)_L \times U(1)_Y$ gauge symmetries of the Standard Model are part of a larger group, such as $SU(5)$ or $SU(10)$ (see Sec. 12.3). If the Standard Model is, in fact, embedded within a GUT, then the couplings g_1, g_2, and g_3 must evolve according to the renormalization group equations to a common value at the scale of GUT symmetry breaking. While this is not possible in the Standard Model, the three gauge couplings do approximately converge to a common value if supersymmetry is included in the calculation (with superpartners near the TeV scale) [125].
- **Connections with gravity**: Supersymmetry seems to be an essential feature of string theory and other related theories of quantum gravity. In particular, if supersymmetry is a local symmetry of nature, it automatically leads to what we would recognize as general relativity, and thus to the existence of gravitational interactions.
- **Dark matter**: In some models, the lightest supersymmetric particle is stable and weakly interacting, making it a potentially viable candidate for dark matter.

For these and other reasons, supersymmetry has long been of great interest to particle physicists. It should be remembered, however, that none of the superpartner particles have been detected to date, and no compelling empirical evidence for supersymmetry has been identified. If the superpartners do exist, they have proven to be more difficult to detect and study than most of the proponents of supersymmetry had once anticipated.

17.1.1 Supersymmetry and the hierarchy problem

From among the particles described by the Standard Model, the Higgs boson provides a unique challenge. In particular, being a scalar, the Higgs boson receives large contributions to its mass through interactions with other Standard Model fields. For example, the production of a virtual fermion-antifermion pair (as shown in the left diagram of Figure 17.1) leads to the following radiative correction to the mass of the Higgs boson:

$$\delta m_h^2 = -\frac{n_c |\lambda_f|^2}{8\pi^2}\Lambda^2 + \cdots, \tag{17.1}$$

where n_c is the number of colors of the fermion ($n_c = 1$ for leptons and $n_c = 3$ for quarks), λ_f is the fermion's coupling to the Higgs (such that the Lagrangian includes a term of the form

[1] In other words, supersymmetry is the only exception to the Coleman-Mandula theorem [360, 361]

Figure 17.1 Interactions involving a fermion (left) or a scalar (right) that induce radiative corrections to the mass of the Higgs boson.

$\mathcal{L} \supset -\lambda_f h \bar{f} f$), and Λ is the cutoff scale of the theory. Notice that these radiative corrections are quadratically sensitive to the cutoff scale, Λ, which can be interpreted as the energy at which new physics begins to alter the high-energy behavior of the theory. This cutoff could plausibly be associated with the Planck scale, for example, or with the scale of grand unification.

The problem posed by Eq. 17.1 is that this correction should increase the mass of the Higgs boson to an absolutely enormous value. The radiative correction from the top quark (with $\lambda_t \approx 0.94$), for example, would lead to $m_h \sim 4 \times 10^{18}\,\text{GeV} \times (\Lambda/M_{\text{Pl}})$, which is wildly larger than the measured value of $m_h \approx 125\,\text{GeV}$. To see how supersymmetry could potentially solve this problem, consider the radiative corrections to the Higgs mass that result from a scalar loop (as shown in the right diagram of Figure 17.1):

$$\delta m_h^2 = \frac{n_c \lambda_S}{16\pi^2}\Lambda^2 + \cdots, \tag{17.2}$$

where λ_S is the scalar's coupling to the Higgs (such that $\mathcal{L} \supset -\lambda_S |h|^2 |S|^2$). Notice that because of the spin-statistics theorem, the sign of this quadratic contribution is opposite to that given in Eq. 17.1, making it possible for cancellations to arise that could reduce the overall mass of the Higgs. To achieve $m_h \lll \Lambda$, however, the fermionic and bosonic corrections would have to be almost identical in magnitude, requiring an incredible degree of fine-tuning. Supersymmetry, however, changes this picture in a critical way. In particular, the Higgs coupling to a given fermion is guaranteed by supersymmetry to be related to that of its superpartner, $|\lambda_f|^2 = \lambda_S$. Thus, in the presence of supersymmetry, the quadratic contributions to the Higgs mass from a given fermion are exactly cancelled by the contributions that arise from its two scalar superpartners, $\delta m_h^2 \approx -(n_c |\lambda_f|^2 \Lambda^2/8\pi^2) + 2 \times (n_c \lambda_S \Lambda^2/16\pi^2) = 0$. As it turns out, this is not only true for the one-loop diagrams explicitly shown here, but for every order of perturbation theory. This cancellation also occurs for spin-1 particles and their fermionic superpartners. Because the masses of the Standard Model particles and their superpartners are not equal, these cancellations are not perfect, but the surviving terms are merely logarithmic rather than quadratic. In this way, so long as the masses of the superpartners are not too much larger than the electroweak scale, supersymmetry makes it possible for us to understand how the Higgs boson could be so relatively light.

17.1.2 Supersymmetric dark matter

Although many different supersymmetric models have been proposed, all such scenarios include two scalar squarks for each quark, two scalar sleptons for each charged lepton, a scalar sneutrino for each neutrino, and a fermionic gluino for each gluon. To avoid gauge anomalies, the Higgs sector must be extended in supersymmetric models to include a minimum of two Higgs doublets, corresponding to two neutral scalar Higgs bosons (h, H), one pseudoscalar Higgs boson (A), and one charged Higgs scalar ($H^{\pm}$). The superpartners

of the electrically neutral Higgs and gauge bosons mix to form (at least) four fermionic states known as neutralinos, while the superpartners of the charged Higgs and $W^{\pm}$ form a pair of fermionic charginos. Supersymmetric scenarios which contain this minimum of new particle content are collectively known as the minimal supersymmetric standard model, or MSSM [128].

As described so far, there is no reason to think that any of the superpartners would be stable, making them largely irrelevant to cosmology. In many supersymmetric models, however, a discrete symmetry known as R-parity is imposed, which can stabilize the lightest superparticle. The R-parity of a given state is defined as follows:

$$R = (-1)^{3(B-L)+2S}, \tag{17.3}$$

where B and L are the baryon and lepton number of the particle, and S is its spin. A quick exercise will show that $R = 1$ for each of the particles in the Standard Model, while $R = -1$ for each of their superpartners. If R-parity is preserved, then superpartners can only be created or destroyed in pairs, which is allowed since $(-1)^2 = 1$. While heavy superparticles can decay to produce a lighter superparticle along with any number of Standard Model states, R-parity ensures that the lightest supersymmetric particle will be stable. Note that R-parity was not originally introduced with the problem of dark matter in mind, but rather to prohibit certain interactions that would have led to an unacceptably fast rate of proton decay [362, 363, 364, 365]. R-parity also prevents the superpartners from mixing with any of the particles of the Standard Model.

In order to be a viable candidate for dark matter, the lightest supersymmetric particle must be electrically neutral and colorless [366]. In the MSSM, the only options that satisfy these requirements are the lightest neutralino and the lightest sneutrino. Other supersymmetric dark matter candidates can arise in less minimal models, however, including gravitinos [367, 366] and axinos, the superpartners of the graviton and axion, respectively. For a variety of phenomenological reasons, neutralinos have long been the most widely studied class of supersymmetric dark matter candidates [368, 369, 370, 371].

17.1.3 The lightest neutralino as a dark matter candidate

The four neutralinos in the MSSM are mixtures of the superpartners of the photon, the Z, and two neutral Higgs bosons. More frequently, this is presented in terms of the gauge eigenstates of these particles, in which the neutralinos are mixtures of the superpartners of the hypercharge gauge boson (the bino, $\tilde{B}$), the neutral W (the wino, $\tilde{W}^3$), and two neutral Higgs bosons (the neutral higgsinos, $\tilde{H}_1^0$, $\tilde{H}_2^0$). The mass eigenstates of the four neutralinos are related to these gauge eigenstates through the following Hermitian mass matrix:

$$\mathcal{M}_{\chi^0} = \begin{pmatrix} M_1 & 0 & -m_Z \cos\beta \sin\theta_W & m_Z \sin\beta \sin\theta_W \\ 0 & M_2 & m_Z \cos\beta \cos\theta_W & -m_Z \sin\beta \cos\theta_W \\ -m_Z \cos\beta \sin\theta_W & -m_Z \cos\beta \cos\theta_W & 0 & -\mu \\ m_Z \sin\beta \sin\theta_W & -m_Z \sin\beta \cos\theta_W & -\mu & 0 \end{pmatrix}, \tag{17.4}$$

where M_1, M_2, and μ are mass parameters associated with the bino, wino, and higgsinos, and $\tan\beta$ is the ratio of the vacuum expectation values of the two Higgs doublets. This matrix

can be diagonalized by a unitary matrix, N, such that

$$N^* \mathcal{M}_{\chi^0} N^\dagger = \begin{pmatrix} m_{\chi_1^0} & 0 & 0 & 0 \\ 0 & m_{\chi_2^0} & 0 & 0 \\ 0 & 0 & m_{\chi_3^0} & 0 \\ 0 & 0 & 0 & m_{\chi_4^0} \end{pmatrix}, \tag{17.5}$$

where, up to their signs, $m_{\chi_i^0}$ are the physical masses of the four neutralinos. N is a 4×4 matrix whose columns are the eigenvectors of $\mathcal{M}_{\chi^0}$. Although the neutralino mass matrix can be diagonalized analytically, the resulting formulas are long and not particularly enlightening. More often, this matrix is diagonalized numerically.

The couplings of the lightest neutralino depend on its composition, which is often expressed in terms of how "bino-like," "wino-like," and "higgsino-like" it is. If $|M_1 \pm \mu|$, $|M_2 \pm \mu| \gg m_Z$, then the lightest neutralino will be either mostly bino-, wino- or higgsino-like, with relatively little mixing between these states. Even in this case, however, the two higgsinos strongly mix with each other (because of the off-diagonal $-\mu$ entries in the mass matrix). In this limit, the masses of the bino-like, wino-like, and higgsino-like neutralinos are approximately given by [372][2]

$$\begin{aligned} m_{\chi^0} &\approx M_1 - \frac{m_Z^2 \sin^2\theta_W (M_1 + \mu \sin 2\beta)}{\mu^2 - M_1^2}, & \chi^0 &\approx \tilde{B} \\ m_{\chi^0} &\approx M_2 - \frac{m_W^2 (M_2 + \mu \sin 2\beta)}{\mu^2 - M_2^2}, & \chi^0 &\approx \tilde{W}^3 \\ m_{\chi^0} &\approx |\mu| + \frac{m_Z^2 ([\mu/|\mu|] \mp \sin 2\beta)(\mu \pm M_1 \cos^2\theta_W \pm M_2 \sin^2\theta_W)}{2(\mu \pm M_1)(\mu \pm M_2)}, & \chi^0 &\approx \tilde{H}_1^0, \tilde{H}_2^0. \end{aligned} \tag{17.6}$$

In contrast, if either $|M_1 \pm \mu|$ or $|M_2 \pm \mu|$ is not too far above the electroweak scale, the bino, wino, and higgsinos will mix with each other more significantly, potentially making the lightest neutralino a highly mixed state.

In order for the four neutralinos to have the same number of degrees of freedom as their neutral gauge and Higgs boson counterparts, these particles must each be Majorana fermions. This has important implications for their annihilations. To see why, consider a pair of neutralinos annihilating into a pair of light Standard Model fermions, $\chi\chi \to f\bar{f}$. In the center of mass frame, these light fermions must be in state with $S = 1$ and $J = 1$, since as fermions they each carry angular momentum. Prior to their annihilation, the neutralinos are in a state of either $S = 0$ and $L = 0$ (and thus $J = 0$) or $S = 1$ and $L = 1$ (and thus $J = 0$, 1, or 2). In the former (s-wave) case, there is no overlap between the initial and final state wavefunctions, suppressing the amplitude for any such processes. In the later case, the amplitude is p-wave in nature, and thus the annihilation cross section scales as $\sigma v \propto v^2 + \mathcal{O}(v^4)$ (see Eq. 9.21). While s-wave amplitudes for neutralino annihilation into fermions are possible, they require a mass insertion and are thus proportional to m_f, leading

[2] In this section, I've adopted a convention in which M_1 and M_2 are necessarily positive, while the sign of μ is taken to be a free parameter.

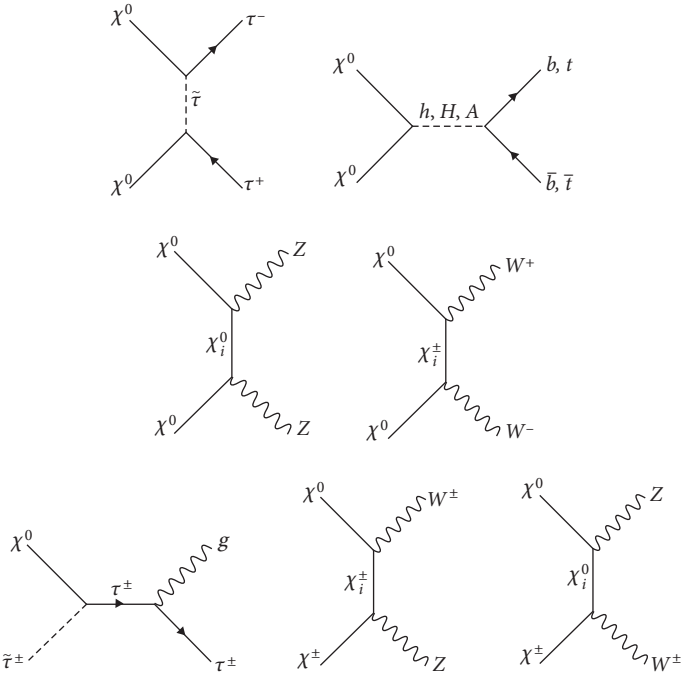


Figure 17.2 Some of the diagrams for neutralino annihilation (top and middle rows) and coannihilation (bottom row) that can play an important role in determining their thermal relic abundance.

to $\sigma v \propto m_f^2$. As a consequence, neutralinos often annihilate preferentially at low velocities into the heaviest kinematically accessible fermionic state, such as $b\bar{b}$ or $t\bar{t}$.

As described above, the couplings of the lightest neutralino depend on its composition. Although the ways in which a neutralino could potentially interact are too numerous to discuss here in any detail, a few general statements can be made. First, in the case of a pure or nearly pure bino-like neutralino, the primary annihilation channel is typically into a pair of right-handed taus through the t-channel exchange of a stau (as shown in the upper left diagram of Figure 17.2). The staus/taus are preferred over other particles in this case due to the large value of their hypercharge, and to the fact that $m_\tau \gg m_\mu, m_e$. Neglecting the effects of mixing between the left- and right-handed staus, and taking $m_\chi \gg m_\tau$, the cross section for this process is given by

$$\sigma v \approx \frac{g_1^4 \, (Y_{\tau_L}^4 + Y_{\tau_R}^4) \, m_{\chi^0}^2 \, (m_{\tilde{\tau}}^2 + m_{\chi^0}^2)}{48\pi \, m_{\tilde{\tau}}^2 \, (m_{\tilde{\tau}} + m_{\chi^0})^4} \, v^2, \tag{17.7}$$

leading to the following thermally averaged value (see Sec. 9.5.3):

$$\begin{aligned} \langle \sigma v \rangle &\approx \frac{g_1^4 \, (Y_{\tau_L}^4 + Y_{\tau_R}^4) \, m_{\chi^0}^2 \, (m_{\tilde{\tau}}^2 + m_{\chi^0}^2)}{8\pi \, m_{\tilde{\tau}}^2 \, (m_{\tilde{\tau}} + m_{\chi^0})^4} \, \frac{m_{\chi^0}}{T} \\ &\sim 5 \times 10^{-27} \, \mathrm{cm}^3/\mathrm{s} \times \left(\frac{m_{\chi^0}}{100 \, \mathrm{GeV}} \right)^2 \left(\frac{200 \, \mathrm{GeV}}{m_{\tilde{\tau}}} \right)^4 \left(\frac{m_{\chi^0}/T}{20} \right), \end{aligned} \tag{17.8}$$

where in the last line I have taken $m_{\chi^0} \ll m_{\tilde{\tau}}$. Recall from Sec. 9.5.2 that in order for a dark matter candidate to freeze out with an acceptable thermal relic abundance, it must annihilate with a cross section of $\langle \sigma v \rangle \sim 2.2 \times 10^{-26}$ cm^3/s. Comparing this with the expression above, we find that the annihilation cross section for a bino-like neutralino will be too small unless the stau mediating the interaction is very light, leading to tension with constraints from colliders.

To avoid overproducing bino-like neutralinos in the early universe, we could instead rely on coannihilations between the lightest neutralino and the staus. In particular, the process $\chi \tilde{\tau} \to \tau \gamma$ (Figure 17.2, lower left) could potentially have depleted the neutralino abundance to an acceptable level in the early universe if enough staus were present at the time of freeze out [373]. As discussed in Sec. 9.5.5, this would only have been possible if the mass of the coannihilating stau is no more than roughly 10% greater than that of the lightest neutralino.

If the lightest neutralino is mostly bino, but has a nonnegligible higgsino component, it will also couple significantly to the neutral Higgs bosons, allowing for efficient s-channel annihilation, $\chi\chi \to h, H, A \to f\bar{f}$ (Figure 17.2, upper right). While the annihilation amplitudes that result from h and H exchange are each p-wave, and thus are suppressed at low velocities, the amplitude from the exchange of the pseudoscalar Higgs, A, is s-wave, providing a potentially large annihilation rate. The cross section for $\chi^0\chi^0 \to f\bar{f}$ through A-exchange is given by

$$\sigma v \approx \frac{3\lambda_{\chi\chi A}^2 \lambda_{ffA}^2 m_\chi^2 \sqrt{1 - m_f^2/m_\chi^2}}{2\pi[(s - m_A^2)^2 + m_A^2 \Gamma_A^2]}, \tag{17.9}$$

where $s \approx m_\chi^2(4 + v^2)$ and the coupling of the neutralino to the pseudoscalar Higgs is given by

$$\lambda_{\chi\chi A} = g_2 \tan\theta_W N_{\tilde{B}} \big[\sin\beta N_{\tilde{H}_1^0} - \cos\beta N_{\tilde{H}_2^0} \big]. \tag{17.10}$$

The quantities $N_{\tilde{B}}$, $N_{\tilde{H}_1^0}$ and $N_{\tilde{H}_2^0}$ in this expression describe the composition of the lightest neutralino, such that $\chi^0 = N_{\tilde{B}}\tilde{B} + N_{\tilde{W}^3}\tilde{W}^3 + N_{\tilde{H}_1^0}\tilde{H}_1^0 + N_{\tilde{H}_2^0}\tilde{H}_2^0$. The couplings of the pseudoscalar Higgs to Standard Model fermions are given by

$$\lambda_{ffA} = \frac{-g_2 m_f \tan\beta}{2m_W}, \qquad f = e, \mu, \tau, d, s, b, \tag{17.11}$$

$$\lambda_{ffA} = \frac{-g_2 m_f \cot\beta}{2m_W}, \qquad f = u, c, t.$$

Notice that while these Yukawa couplings are proportional to the fermion masses, they are also enhanced by a factor of $\tan\beta$ for down-type fermions, and suppressed by a factor of $\cot\beta$ for up-type fermions. In terms of these couplings, the width of the pseudoscalar Higgs is given by

$$\Gamma_A = \sum_f \frac{n_c m_A \lambda_{ffA}}{8\pi} \left(1 - \frac{4m_f^2}{m_A^2}\right)^{1/2}. \tag{17.12}$$

For relatively large values of $\tan\beta$, annihilations in this case proceed mostly to $b\bar{b}$. Far away from the resonance, $|4m_\chi^2 - m_A^2| \gtrsim m_A \Gamma_A$, the cross section for this process is approximately

given by

$$\sigma v \approx 2 \times 10^{-26}\,\mathrm{cm}^3/\mathrm{s} \times \left(\frac{\tan\beta}{50}\right)^2 \left(\frac{m_\chi}{100\,\mathrm{GeV}}\right)^2 \left(\frac{N_{\tilde{B}}^2}{0.9}\right)\left(\frac{N_{\tilde{H}_1^0}^2}{0.1}\right)\left(\frac{(400\,\mathrm{GeV})^2}{m_A^2 - 4m_\chi^2}\right)^2, \tag{17.13}$$

which is large enough to potentially obtain an acceptable thermal relic abundance. In this case, however, the lightest neutralino will also have a large coupling to the scalar Higgs bosons (h, H), leading to a sizable cross section for the elastic scattering of neutralinos with nucleons, and thus to significant tension with constraints from direct detection experiments (see Chapter 18). To evade this issue, one could consider neutralinos that can annihilate through the pseudoscalar Higgs resonance, $|s - m_A^2| \lesssim m_A \Gamma_A$. In this on-resonance case, the annihilation cross section is given by

$$\begin{aligned}
\sigma v &\approx \frac{3\lambda_{\chi\chi A}^2 \lambda_{ffA}^2 m_\chi^2 \sqrt{1 - m_f^2/m_\chi^2}}{2\pi m_A^2 \Gamma_A^2}, \\
&\approx \frac{32\pi g_2^2 \tan^2\theta_W m_\chi^2 N_{\tilde{B}}^2 N_{\tilde{H}_1^0}^2}{3m_A^4} \\
&\approx 3 \times 10^{-26}\,\mathrm{cm}^3/\mathrm{s} \times \left(\frac{m_\chi}{300\,\mathrm{GeV}}\right)^2 \left(\frac{600\,\mathrm{GeV}}{m_A}\right)^4 \left(\frac{N_{\tilde{B}}^2}{0.999}\right)\left(\frac{N_{\tilde{H}_1^0}^2}{0.001}\right),
\end{aligned} \tag{17.14}$$

where in the last two lines I have taken $\tan\beta \gg 1$. After computing a thermal average of this cross section as described in Sec. 9.5.4, one finds that this annihilation channel could provide an acceptable relic abundance of neutralino dark matter, without necessarily conflicting with direct detection constraints.

For both wino-like ($N_{\tilde{W}^3}^2 \approx 1$) and higgsino-like ($N_{\tilde{H}_1^0}^2 + N_{\tilde{H}_2^0}^2 \approx 1$) neutralinos, annihilations can proceed quite efficiently to ZZ through the t-channel exchange of any of the four neutralinos, or to W^+W^- through the t-channel exchange of either chargino (as shown in the middle row of Figure 17.2). Furthermore, in each of these cases, the mass of the lightest neutralino is only slightly lower than that of the lightest chargino, making coannihilations between these states potentially important. To understand why this is the case, consider the chargino mass matrix:

$$\mathcal{M}_{\chi^\pm} = \begin{pmatrix} M_2 & \sqrt{2}\, m_W \sin\beta \\ \sqrt{2}\, m_W \cos\beta & \mu \end{pmatrix}. \tag{17.15}$$

The square of this matrix, $\mathcal{M}_{\chi^\pm}^\dagger \mathcal{M}_{\chi^\pm}$, has the following eigenstates:

$$\begin{aligned}
m_{\chi_1^\pm, \chi_2^\pm}^2 = \frac{1}{2}\Big[& M_2^2 + \mu^2 + 2m_W^2 \\
& \mp \sqrt{(M_2^2 + \mu^2 + 2m_W^2)^2 - 4(\mu M_2 - m_W^2 \sin 2\beta)^2} \Big].
\end{aligned} \tag{17.16}$$

In the $|M_2 \pm \mu| \gg m_W$ limit, the chargino masses are approximately given by [372]

$$m_{\chi^\pm} \approx M_2 - \frac{m_W^2 (M_2 + \mu \sin 2\beta)}{\mu^2 - M_2^2}, \qquad \chi^\pm \approx \tilde{W}^\pm, \tag{17.17}$$

$$m_{\chi^\pm} \approx |\mu| + \left(\frac{\mu}{|\mu|}\right) \frac{m_W^2 (\mu + M_2 \sin 2\beta)}{\mu^2 - M_2^2}, \qquad \chi^\pm \approx \tilde{H}^\pm.$$

In the limit in which the lightest neutralino is highly wino-like ($M_2 \ll |\mu|, M_1$), both this state and the wino-like chargino will have a mass of $m_{\chi^0} \approx m_{\chi^\pm} \approx M_2$. Similarly, if the lightest neutralino is higgsino-like ($|\mu| \ll M_1, M_2$), there will be a near degeneracy between the two lightest neutralinos and the lightest chargino, $m_{\chi_1^0} \approx m_{\chi_2^0} \approx m_{\chi^\pm} \approx |\mu|$. While the tree-level masses of these states can be found by diagonalizing the respective mass matrices, there will also be small loop-level corrections that increase the mass splitting between these states by $\Delta(m_{\chi^\pm} - m_{\chi^0}) \approx 0.17\,\text{GeV}$ in the case of winos, and by $\Delta(m_{\chi^\pm} - m_{\chi^0}) \approx 0.35\,\text{GeV}$ for the higgsinos [374]. Electroweak symmetry breaking also induces a small mass splitting between the two higgsino-like neutralinos, typically on the order of $\sim 0.1 - 10\,\text{GeV}$ [375].

In scenarios in which the lightest neutrino is mostly wino-like, the wino-like charginos will be fully present in the thermal bath during the process of dark matter freeze out, making it possible for coannihilations between these two states to deplete the abundance of these particles (see Sec. 9.5.5). The annihilation and coannihilation of wino-like neutralinos and charginos into pairs of gauge boson through the t-channel exchange of neutralinos and charginos (see Figure 17.2) is very efficient, with cross sections that are approximately given by

$$\sigma v(\chi^0 \chi^0 \to W^+ W^-, ZZ) \approx \frac{g_2^4}{8\pi M_2^2}, \tag{17.18}$$

$$\sigma v(\chi^0 \chi^\pm \to Z W^\pm) \approx \frac{7 g_2^4}{32\pi M_2^2},$$

$$\sigma v(\chi^+ \chi^- \to W^+ W^-, ZZ) \approx \frac{9 g_2^4}{32\pi M_2^2},$$

$$\sigma v(\chi^\pm \chi^\pm \to W^\pm W^\pm) \approx \frac{g_2^4}{16\pi M_2^2},$$

where $m_{\chi^0} \approx m_{\chi^\pm} \approx M_2$. Following the procedure described in Sec. 9.5.5, we can calculate the effective cross section for neutralino and chargino annihilation and coannihilation:

$$\begin{aligned} \langle \sigma_{\text{eff}} v \rangle &\approx \frac{1}{9} \sigma v(\chi^0 \chi^0) + \frac{4}{9} \sigma v(\chi^0 \chi^\pm) + \frac{2}{9} \sigma v(\chi^+ \chi^-) + \frac{2}{9} \sigma v(\chi^\pm \chi^\pm) \\ &\approx \frac{3 g_2^4}{16\pi M_2^2} \\ &\approx 2.2 \times 10^{-26}\,\text{cm}^3/\text{s} \times \left(\frac{2.4\,\text{TeV}}{M_2}\right)^2. \end{aligned} \tag{17.19}$$

Taking into account the non-perturbative effects known as Sommerfeld enhancements [376, 377], one finds that a pure wino freezes out with a thermal relic abundance that is

equal to the measured density of dark matter for a mass of $m_{\chi^0} \sim 3\,\text{TeV}$ [378, 379, 380]. Importantly, the wino's scattering cross section with nuclei is generated only through loop-level diagrams, allowing this dark matter candidate to evade the constraints from current direct detection experiments [381, 382].

The case of a higgsino-like lightest neutralino is similar, but also includes a nearly degenerate second neutralino. In this scenario, χ_1^0, χ_2^0, and $\chi^\pm$ are all fully present in the thermal bath at the time of dark matter freeze out. The cross sections for the relevant annihilation and coannihilation processes in this case are approximately given as follows:

$$\sigma v(\chi_1^0\chi_1^0 \to W^+W^-, ZZ) \approx \frac{g_2^4}{128\pi\mu^2}\left[\frac{3}{2} + \tan^2\theta_W + \frac{\tan^4\theta_W}{2}\right], \tag{17.20}$$

$$\sigma v(\chi_2^0\chi_2^0 \to W^+W^-, ZZ) \approx \frac{g_2^4}{128\pi\mu^2}\left[\frac{3}{2} + \tan^2\theta_W + \frac{\tan^4\theta_W}{2}\right],$$

$$\sigma v(\chi_1^0\chi_2^0 \to W^+W^-, ZZ) \approx \frac{g_2^4}{128\pi\mu^2}\left[\frac{29}{4} + \frac{21}{2}\tan^4\theta_W\right],$$

$$\sigma v(\chi_1^0\chi^\pm \to W^+W^-, ZZ) \approx \frac{g_2^4}{128\pi\mu^2}\left[6 + \tan^2\theta_W\right],$$

$$\sigma v(\chi_2^0\chi^\pm \to W^+W^-, ZZ) \approx \frac{g_2^4}{128\pi\mu^2}\left[6 + \tan^2\theta_W\right],$$

$$\sigma v(\chi^\pm\chi^\mp \to W^+W^-, ZZ) \approx \frac{g_2^4}{128\pi\mu^2}\left[\frac{29}{4} + \tan^2\theta_W + 11\tan^4\theta_W\right],$$

$$\sigma v(\chi^\pm\chi^\pm \to W^+W^-, ZZ) \approx \frac{g_2^4}{64\pi\mu^2}.$$

Combining these, we obtain the effective cross section for higgsino annihilation and coannihilation:

$$\begin{aligned}\langle\sigma_{\text{eff}}\, v\rangle &\approx \frac{1}{16}\sigma v(\chi_1^0\chi_1^0) + \frac{1}{16}\sigma v(\chi_2^0\chi_2^0) + \frac{1}{8}\sigma v(\chi_1^0\chi_2^0) + \frac{1}{4}\sigma v(\chi_1^0\chi^\pm) \\ &\quad + \frac{1}{4}\sigma v(\chi_2^0\chi^\pm) + \frac{1}{8}\sigma v(\chi^\pm\chi^\mp) + \frac{1}{16}\sigma v(\chi^\pm\chi^\pm) \\ &\approx \frac{g_2^4}{512\pi\mu^2}\left[\frac{41}{2} + \frac{49}{8}\tan^2\theta_W + 11\tan^4\theta_W\right] \\ &\approx 2.2\times10^{-26}\,\text{cm}^3/\text{s} \times \left(\frac{1.1\,\text{TeV}}{\mu}\right)^2.\end{aligned} \tag{17.21}$$

Although higgsino-like neutralinos can scatter with nuclei through the exchange of a Z, this process requires the incoming neutralino to be converted into the (slightly) heavier neutral higgsino. As a result, this process can be kinematically suppressed if the incoming particle does not have enough kinetic energy to scatter inelastically [375].

A decade or more ago, most of the parameter space of supersymmetric models remained potentially viable. The most restrictive constraints at the time were based on measurements

from the Large Electron-Positron Collider (LEP), which required any electrically charged superpartners to be heavier than approximately $\sim 10^2$ GeV. Since that time, however, the null results of searches for supersymmetric particles at the Large Hadron Collider (LHC) have shown that any gluinos, or first or second generation squarks, must be heavier than a couple of TeV. Similarly, the masses of any charginos or charged sleptons must be greater than $\sim 200 - 300$ GeV or so. Over roughly the same period of time, direct detection experiments (see Chapter 18), attempting to detect the scattering of dark matter particles with nuclei, have become far more sensitive, ruling out many models of neutralino dark matter.

In light of these combined constraints, much of the supersymmetric parameter space now seems to be excluded. In response to this situation, attention is now often focused toward those specific regions of parameter space that have not yet been fully explored. Some of the most prominent of these remaining regions include the following:

- Coannihilation regions: In scenarios in which the lightest neutralino is approximately degenerate in mass with another state, an acceptable thermal relic abundance can sometimes be obtained without inducing a large scattering cross section with nuclei.
- The A-funnel region: In models in which $m_{\chi^0} \sim m_A/2$, the neutralino can often annihilate very efficiently through the pseudoscalar Higgs resonance, leading to an acceptable thermal relic abundance without a large scattering cross section with nuclei.
- Nearly pure winos: The large couplings of winos to gauge bosons allow them to freeze out with an acceptable abundance, while scattering with nuclei only through loop diagrams.
- Nearly pure higgsinos: The large couplings of higginos to gauge bosons allow them to freeze out with an acceptable abundance. If the mass splitting between the two higgsino-like neutralinos is large enough to suppress their inelastic scattering with nuclei, they can evade existing constraints.
- Direct detection blind spots: For particular ratios of μ and m_{χ^0} (which depend on the values of $\tan\beta$ and m_A), cancellations can occur that suppress the lightest neutralino's scattering cross section with nuclei, while still allowing for an acceptable thermal relic abundance [383, 384].

Although enthusiasm for supersymmetry has fallen well below its historical peak, these regions of potentially viable parameter space have enabled the lightest neutralino to remain a somewhat popular candidate for the dark matter of our universe.

17.2 Sterile neutrinos

The three known neutrino species are too light and interact too frequently with other particles to constitute good candidates for dark matter. There could, however, exist other stable (or cosmologically long-lived) fermions that are heavier and that do not experience the full strength of the weak interaction. Furthermore, such particles could mix with the Standard Model neutrinos, providing a novel mechanism for them to be created in the early universe.

As discussed in Section 11.1, the three neutrinos are predicted to be exactly massless in the Standard Model. Whereas the Higgs mechanism is capable of generating masses for

the quarks and charged leptons, it does this through interactions involving both left- and right-handed fermions. Because the Standard Model neutrinos are uniquely left-handed (and the antineutrinos are uniquely right-handed), no neutrino masses can be generated in this way. Since the 1990s, however, it has been known that the Standard Model neutrinos can transform, or oscillate, from one flavor into another. This, in turn, implies that the masses of the three neutrinos must be (slightly) different from one another. More specifically, neutrino oscillations occur because the flavor eigenstates of the three neutrinos, ν_e, ν_μ, and ν_τ, are each mixtures of the three mass eigenstates, ν_1, ν_2, and ν_3. How the neutrino masses are generated remains one of the most important outstanding puzzles in all of particle physics.

Although the neutrinos are massless in the Standard Model, there are natural extensions of this theory in which small masses for these particles can be generated through variations of what is known as the seesaw mechanism [385, 386, 107, 106, 387]. Such scenarios predict the existence of sterile neutrinos, which do not interact through any of the Standard Model gauge forces. The mass matrix that describes one such sterile neutrino and a single active neutrino can be written as follows (see Eq. 11.6):

$$\mathcal{M}_\nu = \begin{pmatrix} 0 & \mu \\ \mu & M \end{pmatrix}, \tag{17.22}$$

where μ is the Dirac mass and M is the right-handed Majorana mass (see Eqs. 11.4 and 11.5). This matrix has two eigenvalues, corresponding to

$$m_{\nu_{1,2}} = \left| \frac{M \pm \sqrt{M^2 + 4\mu^2}}{2} \right|. \tag{17.23}$$

In the $\mu \ll M$ limit, this corresponds to a mostly active state with a very small mass, $m_{\nu_1} \equiv m_a \approx \mu^2/M$, and a much heavier, mostly sterile state with $m_{\nu_2} \equiv m_s \approx M$. These two mass eigenstates can be described as mixtures of the active and sterile flavor eigenstates:

$$\begin{aligned} |\nu_1\rangle &= \cos\theta\, |\nu_a\rangle - \sin\theta\, |\nu_s\rangle, \\ |\nu_2\rangle &= \sin\theta\, |\nu_a\rangle + \cos\theta\, |\nu_s\rangle, \end{aligned} \tag{17.24}$$

where $\theta \sim \mu/M$ is the mixing angle between the active and sterile states. As a result of this mixing, the sterile neutrinos have an effective interaction that functions exactly as the weak interaction, but that is suppressed by a factor of $\sin 2\theta$. Scenarios that involve a mass matrix of the form shown in Eq. 17.22 can naturally explain why the active neutrinos are so light, and generally feature a large hierarchy between the masses of the sterile and active states.

17.2.1 Sterile neutrino production through oscillations

If the degree of mixing between the sterile and active neutrinos is very small, the sterile neutrinos will not reach thermal equilibrium with the Standard Model bath in the early universe. As pointed out by Dodelson and Widrow [388], however, even a very small degree of mixing between these states can lead to a significant population of sterile neutrinos being generated through the oscillations of active neutrinos followed by their collisions with particles in the Standard Model bath. In this class of scenarios, the production of sterile neutrinos can be thought of as a variant of thermal freeze in, as discussed in Sec. 9.7.1.

The rate of sterile neutrino production in the early universe depends on both the scattering rate of the active neutrinos, and on their mixing angle with the sterile state. The quantity θ, as described above, represents the active-sterile mixing angle in vacuum.

If this mixing occurred in the early universe at a level set by this angle, the production of sterile neutrinos would be dominated by the hottest era of our universe's history, making it an example of UV-dominated freeze in. Thermal effects arising from interactions with the Standard Model plasma, however, induce a potential that shifts the mixing angle to the following effective value (see Eq. 11.29):

$$\sin^2 2\theta_{\rm eff} = \frac{\sin^2 2\theta}{\sin^2 2\theta + [\cos 2\theta - (2E_\nu/\Delta m^2)V]^2}, \tag{17.25}$$

where E_ν is the energy of the neutrino and $\Delta m^2 \approx m_s^2$ is the difference between the squared masses of the sterile and active states. In the absence of any large matter-antimatter asymmetries in the thermal bath, the potential can be written as [389]:

$$V = -\frac{64\pi\,\Gamma}{15(g_1^2+g_2^2)}, \tag{17.26}$$

where Γ is the collision rate of the neutrinos in the thermal bath (we will consider scenarios with a significant matter-antimatter asymmetry later in this section). For tau neutrinos at temperatures $m_\mu \gg T \gg m_e$, for example, this collision rate is approximately given by

$$\begin{aligned}
\Gamma &= \sum_i \sigma_{\nu_\tau i}\, n_i \qquad (17.27)\\
&= \sigma_{\nu_\tau \nu_e} n_{\nu_e} + \sigma_{\nu_\tau \nu_\mu} n_{\nu_\mu} + \sigma_{\nu_\tau \nu_\tau} n_{\nu_\tau} + \sigma_{\nu_\tau \bar{\nu}_e} n_{\bar{\nu}_e} + \sigma_{\nu_\tau \bar{\nu}_\mu} n_{\bar{\nu}_\mu} + \sigma_{\nu_\tau \bar{\nu}_\tau} n_{\bar{\nu}_\tau} + \sigma_{\nu_\tau e^\pm} n_{e^\pm}\\
&\approx \frac{6G_F^2 s}{12\pi}\frac{3\zeta(3)T^3}{4\pi^2} + \frac{6G_F^2 s}{12\pi}\frac{3\zeta(3)T^3}{4\pi^2} + \frac{12G_F^2 s}{12\pi}\frac{3\zeta(3)T^3}{4\pi^2}\\
&\quad + \frac{2G_F^2 s}{12\pi}\frac{3\zeta(3)T^3}{4\pi^2} + \frac{2G_F^2 s}{12\pi}\frac{3\zeta(3)T^3}{4\pi^2} + \frac{[9+(1-4\sin^2\theta_W)^2]G_F^2 s}{12\pi}\frac{3\zeta(3)T^3}{4\pi^2}\\
&\quad + 2\frac{[1+(1-4\sin^2\theta_W)^2]G_F^2 s}{12\pi}\frac{3\zeta(3)4T^3}{4\pi^2}\\
&= \frac{45G_F^2\zeta(3)T^3 s}{16\pi^3}\left[1+\frac{(1-4\sin^2\theta_W)^2}{5}\right]\\
&\approx \frac{7\pi\,G_F^2 T^4 E_\nu}{24}\left[1+\frac{(1-4\sin^2\theta_W)^2}{5}\right],
\end{aligned}$$

where $s \approx (8/3)\langle E_{\rm target}\rangle E_\nu \approx (8/3)(3.15\,T)E_\nu \approx 8.4\,TE_\nu$. This calculation can be further generalized to consider neutrinos of all flavors, and at higher temperatures where it is necessary to include scattering with muons and taus [390].

Now that we have established the temperature dependance of the effective sterile-active mixing angle, we can determine how the process of sterile neutrino production will play out in the early universe. The probability that a neutrino that starts out in a purely active state will be found in a sterile state after a time t is given by

$$P_{\nu_a \to \nu_s} = \sin^2 2\theta_{\rm eff}\,\sin^2\left(\frac{vt}{L}\right), \tag{17.28}$$

where v is the velocity of the neutrino, and L is the length scale over which the oscillations occur. We will typically find ourselves in the limit of $vt \gg L$, for which the probability simplifies to the average value of $P_{\nu_a \to \nu_s} \approx (1/2)\sin^2 2\theta_{\rm eff}$. The rate at which an active neutrino will be converted into a sterile neutrino is given by the product of this oscillation probability and the collision rate, Γ.

Now that we have the rate for sterile neutrino production, we can write down the Boltzmann equation for the evolution of their number density:

$$\frac{dn_s}{dt} = -3Hn_s + n_a \frac{\sin^2 2\theta_{\rm eff}}{2}\,\Gamma. \tag{17.29}$$

The first term on the right-hand side of this equation accounts for the dilution from Hubble expansion, while the rightmost term is the production rate through active-sterile oscillations. As written in this equation, the values of $\sin^2 2\theta_{\rm eff}$ and Γ are treated as the same for each neutrino. To obtain a more accurate result, we should integrate over the energy distribution of the neutrinos:

$$\begin{aligned}\frac{dn_s}{dt} &= -3Hn_s + \int \frac{dN_\nu}{dE_\nu}\frac{\sin^2 2\theta_{\rm eff}}{2}\,\Gamma\, dE_\nu \qquad (17.30)\\ &= -3Hn_s + \int \frac{E_\nu^2}{\pi^2(e^{E_\nu/T}+1)} \\ &\quad \times \frac{1}{2}\left(\frac{\sin^2 2\theta}{\sin^2 2\theta + [\cos 2\theta - (2E_\nu/\Delta m^2)V]^2}\right)\left(\frac{-15V(g_1^2+g_2^2)}{64\pi}\right) dE_\nu.\end{aligned}$$

At high temperatures, the thermal potential strongly suppresses the effective mixing angle, and thus the production rate of sterile neutrinos. This continues until $2E_\nu V/m_s^2 \sim 1$, at which point the effective mixing angle becomes approximately constant. In this regime, the production rate scales as $\sim T^8$, dropping rapidly as the universe expands and cools. Numerically integrating this expression, one finds that the proportion of active neutrinos that are transformed into sterile neutrinos per log interval of temperature is maximized near $T \sim 100\,{\rm MeV} \times (m_s/{\rm keV})^{1/3}$, with the bulk of the sterile neutrino production taking place at temperatures near this peak value. In Figure 17.3, I show the evolution of the number density of sterile neutrinos in this scenario, for the case of $m_s = 3\,{\rm keV}$ and with $\sin^2 2\theta = 10^{-8}$. As expected, the abundance rises rapidly up until $T \sim 100\,{\rm MeV}$, after which the production rate rapidly drops and the comoving number density becomes approximately constant.

Numerically, the Boltzmann equation results in the following abundance of sterile neutrinos being produced through the oscillations of electron neutrinos:

$$\rho_{\nu_s} \approx 1.3 \times 10^{-6}\,{\rm GeV/cm^3} \times \left(\frac{m_s}{\rm keV}\right)^2 \left(\frac{\sin^2 2\theta}{7\times 10^{-8}}\right)^{1.23}. \tag{17.31}$$

A number of constraints can be placed on sterile neutrinos with masses and mixing angles in this range. First of all, being fermions, they are subject to the Pauli exclusion principle, allowing us to apply the Tremaine-Gunn bound [65] of $m_s \gtrsim 1\,{\rm keV}$ (see Sec. 9.4). Sterile neutrinos with masses not too far above the Tremaine-Gunn bound can also be constrained through their impact on large-scale structure. As discussed in Sec. 10.6, thermal relics in the keV-scale mass range will act as warm dark matter, leading to the suppression of power on scales smaller than $k \sim 10\,{\rm Mpc}^{-1}$. Based on these effects, measurements of the

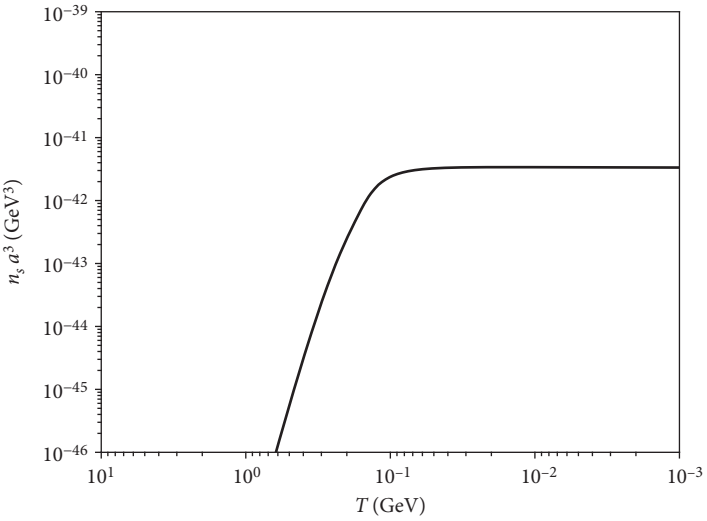


Figure 17.3 The comoving number density of a 1 keV sterile neutrino with a mixing angle of $\sin^2 2\theta = 7 \times 10^{-8}$ as a function of temperature.

Lyman-alpha forest, and of the number of Milky Way satellite galaxies, rule out any such dark matter candidates lighter than ~ 5 keV [102, 103, 104]. In the case of sterile neutrinos that are produced through oscillations, however, the resulting momentum distribution is somewhat different—and somewhat hotter—than that predicted for an ordinary thermal relic [391, 392, 393, 91]. As a result, these same measurements rule out sterile neutrinos lighter than approximately $\sim 20 - 30$ keV if generated through the Dodelson-Widrow mechanism.

Constraints on sterile neutrino dark matter have also been derived from their decays. The main decay mode for a keV-scale sterile neutrino occurs through a tree-level diagram to three active neutrinos, as shown in the leftmost diagram of Figure 17.4. The decay width for this process is given by [394]

$$\Gamma_{\nu_s \to \nu\nu\bar{\nu}} = \frac{G_F^2 m_s^5}{918\pi^3} \sin^2 2\theta, \tag{17.32}$$

leading to a lifetime of

$$\tau_s = \Gamma^{-1}_{\nu_s \to \nu\nu\bar{\nu}} \approx 2 \times 10^{27}\,\text{s} \times \left(\frac{\text{keV}}{m_s}\right)^5 \left(\frac{7 \times 10^{-8}}{\sin^2 2\theta}\right). \tag{17.33}$$

Measurements of the CMB indicate that no more than a small fraction of the dark matter has decayed over cosmic history, allowing us to constrain $\tau_s \gtrsim 160\,\text{Gyr} \approx 5 \times 10^{18}$ s [395]. From this, we can place an upper limit on m_s for a given value of $\sin^2 2\theta$. A more restrictive constraint, however, comes from the loop-induced decays of sterile neutrinos to an active neutrino and a photon, $\nu_s \to \nu\gamma$, as shown in the center and right diagrams of Figure 17.4. The width for this process is given by [17]

$$\Gamma_{\nu_s \to \nu\gamma} = \frac{9 G_F^2 \alpha m_s^5}{2048\pi^4} \sin^2 2\theta. \tag{17.34}$$

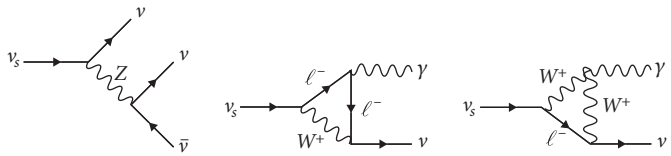


Figure 17.4 Tree-level (left) and loop-level (center, right) diagrams for the decays of sterile neutrinos.

Despite the fact that this process is responsible for only $27\alpha/8\pi \approx 0.4\%$ of all sterile neutrino decays, such occurrences produce a mono-energetic photon, $E_\gamma \approx m_s/2$, leading to a distinctive signal that can be searched for very effectively using X-ray telescopes [396, 397, 398].

These constraints are summarized in Figure 17.5, along with the regions of parameter space in which oscillations are predicted to generate an abundance of sterile neutrinos that is equal to the dark matter density. Between X-ray line searches and measurements of the Lyman-alpha forest and Milky Way satellite galaxies, the entirety of this parameter space is excluded in the Dodelson-Widrow case.

17.2.2 Sterile neutrinos in the presence of a lepton asymmetry

In light of the challenges faced by sterile neutrinos produced through the Dodelson-Widrow mechanism, a number of less minimal scenarios have been proposed in which the production rate of these particles could be increased in the early universe, while still having their mixing angles be small enough to evade the bounds from X-ray line searches. Most notably, it was pointed out by Shi and Fuller [401] that the rate of sterile neutrino production could be resonantly enhanced if there was a large asymmetry between the numbers of neutrinos and antineutrinos present in the early universe.

In the calculation of the abundance of sterile neutrinos produced through the Dodelson-Widrow mechanism, we assumed that equal amounts of matter and antimatter were present in the thermal bath. In the presence of a lepton asymmetry, however, the potential and thus the effective mixing angle is modified. Including the effects of such a lepton asymmetry, the potential for the oscillations of neutrinos of flavor α can be written as

$$V = V^T + \sqrt{2} G_f \left[2 L_{\nu_\alpha} n_\gamma + \sum_{\beta \neq \alpha} L_{\nu_\beta} n_\gamma \right], \tag{17.35}$$

where V^T is the thermal potential as given in Eq. 17.26, and the lepton asymmetries are defined as

$$L_{\nu_\alpha} \equiv \frac{n_{\nu_\alpha} - n_{\bar{\nu}_\alpha}}{n_\gamma}. \tag{17.36}$$

Note that I have neglected in these expressions the asymmetries in the charged leptons and baryons, which are known to have been small, $\sim 6 \times 10^{-10}$. If the asymmetries among the neutrinos had been comparably small, they would have had a negligible impact on the production of sterile neutrinos in the early universe. The values of L_{ν_α}, however, are only weakly constrained. In particular, measurements of the primordial helium abundance, the radiation density at the time of photon decoupling, and large-scale structure collectively yield the following constraints: $-0.041 < L_{\nu_e} < 0.79$ and $|L_{\nu_\mu, \nu_\tau}| < 6.0$ [402, 403, 404, 405,

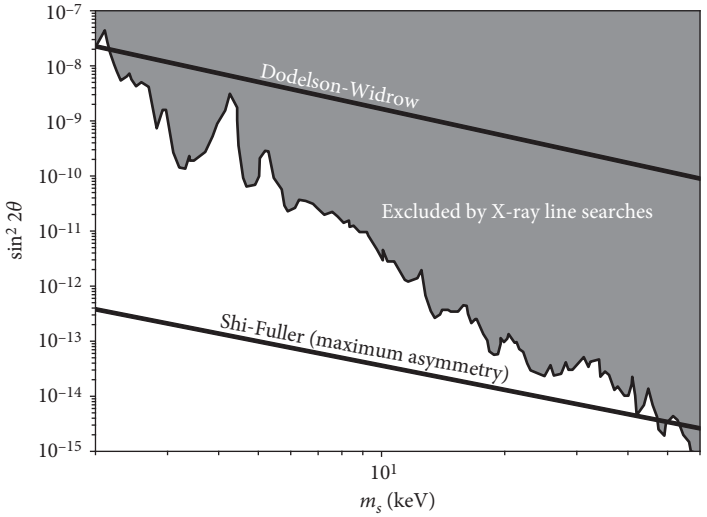


Figure 17.5 Constraints from X-ray line searches [392, 399, 400] on the parameter space of sterile neutrino dark matter, as produced through nonresonant oscillations with active neutrinos (the Dodelson-Widrow mechanism), or through resonant oscillations (the Shi-Fuller mechanism) resulting from the maximum lepton asymmetry allowed by BBN and other cosmological observations. After combining these constraints with those on warm dark matter, the Dodelson-Widrow scenario is entirely ruled out as the primary mechanism for dark matter's production.

406]. In light of this, it remains possible that a very large neutrino asymmetry could have been present in the early universe, many order of magnitude larger than that found among the baryons or charged leptons.

The main idea behind the Shi-Fuller mechanism is that, in the presence of a large lepton number asymmetry, the potential can be large enough to cause the terms in the square bracket in the denominator of Eq. 17.25 to cancel, leading to $\sin^2 2\theta_{\rm eff} \sim 1$ in the presence of the thermal bath. This resonance occurs when

$$\frac{m_s^2}{2E_\nu} \simeq V \approx \sqrt{2} G_f \left[2L_{\nu_\alpha} n_\gamma + \sum_{\beta \neq \alpha} L_{\nu_\beta} n_\gamma \right], \tag{17.37}$$

corresponding to a temperature of

$$T \approx 14\,{\rm MeV} \times \left(\frac{m_s}{\rm keV} \right)^{1/2} \left(\frac{3.15\,T}{E_\nu} \right)^{1/4} \left(2L_{\nu_\alpha} + \sum_{\beta \neq \alpha} L_{\nu_\beta} \right)^{-1/4}. \tag{17.38}$$

The production rate of sterile neutrinos is enhanced at temperatures near this resonance, allowing oscillations to produce the measured dark matter abundance with a smaller mixing angle (in vacuum), potentially evading constraints from X-ray line searches. Furthermore, sterile neutrinos produced through the Shi-Fuller mechanism can, in some cases, be colder than predicted in the Dodelson-Widrow scenario, leading to somewhat weaker constraints from the Lyman-alpha forest and satellite galaxy counts [104, 407, 393].

In Figure 17.5, I show the parameter space in which sterile neutrinos can be resonantly produced to constitute the entire dark matter density, for the case of the maximum lepton

number asymmetry that is currently allowed by BBN and other cosmological observations. In this extreme case, sterile neutrinos could be as heavy as several tens of keV without exceeding the constraints from X-ray line searches.

17.3 Axions

The axion was originally introduced as part of a solution to what is known as the "strong CP problem." This problem basically boils down to the question of why the electric dipole moment of the neutron is so small. In fact, this quantity is so small that it is yet to be measured, despite considerable experimental effort.

Unlike the electromagnetic force, the structures of the strong and weak forces each allow for the violation of charge-parity (CP) symmetry. Technically speaking, this is because they are based on non-abelian gauge symmetries, unlike $U(1)_{\rm EM}$, is which abelian. As already pointed out in Chapter 12, CP violation is, in fact, observed among the weak interactions. It has not, however, been observed in any of the interactions associated with QCD.

The Lagrangian of QCD includes the following term, known as the θ term, which allows for CP violation:

$$\mathcal{L} = -\frac{\theta_{\rm QCD} g_s^2}{32\pi^2} G^{a\mu\nu} \tilde{G}_{a\mu\nu}. \tag{17.39}$$

Here, g_s is the strong coupling constant, $G^{a\mu\nu}$ is the gluon field, and $\tilde{G}^{a\mu\nu}$ is its dual. The quantity $\theta_{\rm QCD}$ is an angle whose value quantifies the degree to which CP violation occurs in the strong interactions. It is important to realize that the structure of QCD doesn't provide us with any reason for why $\theta_{\rm QCD}$ should take on any particular value; all values between π and $-\pi$ appear to be equally possible.

The strong CP problem arises when the value of $\theta_{\rm QCD}$ is related to that of the neutron's electric dipole moment, d_n:

$$d_n \approx 3.6 \times 10^{-16}\, \theta_{\rm QCD}\, e\,{\rm cm}, \tag{17.40}$$

where e is the charge of the electron. Since laboratory experiments have found $|d_n| < 3 \times 10^{-26}\, e\,{\rm cm}$ [408], this requires $|\theta_{\rm QCD}| \lesssim 10^{-10}$. Given that $\theta_{\rm QCD}$ could have had any value between π and $-\pi$, it is hard to understand why its magnitude would be so tiny. To accommodate this observation, one might posit the existence of some unknown symmetry that forces $\theta_{\rm QCD}$ to be equal to zero. This, however, turns out to be difficult to arrange. In particular, given that $\theta_{\rm QCD}$ receives contributions from the structure of QCD itself and from the values of the quark masses, the latter of which is necessarily finite, the idea that these two contributions would almost exactly cancel one another seems extremely unlikely.

The strong CP problem has motivated a number of extensions of the Standard Model, the most well-known example of which was proposed by Roberto Peccei and Helen Quinn in 1977 [409]. In the Peccei-Quinn solution, the quantity $\theta_{\rm QCD}$ is taken to be a dynamical variable, instead of a mere constant. More specifically, they accomplished this by introducing a new global $U(1)_{\rm PQ}$ symmetry that is broken by a complex scalar field. The potential for this complex scalar field, φ, is given by:

$$V(\varphi) = \lambda \left(|\varphi|^2 - \frac{f_{\rm PQ}^2}{2} \right)^2, \tag{17.41}$$

where f_{PQ} is a parameter that represents the energy scale at which the Peccei-Quinn symmetry is broken and that is sometimes called the axion decay constant. This "wine bottle" or "Mexican hat" potential (see Figure 12.5) has a local maximum at $\varphi = 0$ and a continuous range of minima at $\varphi = (f_{PQ}/\sqrt{2})\, e^{i\phi/f_{PQ}}$, where ϕ/f_{PQ} is the complex phase of φ. Since the minima occur at $\varphi \neq 0$, this field has a vacuum expectation value. Although the magnitude of φ is $f_{PQ}/\sqrt{2}$ in the ground state, all values of the phase, ϕ, are equally energetically preferred; i.e., $V = 0$ for all $\varphi = (f_{PQ}/\sqrt{2})\, e^{i\phi/f_{PQ}}$, regardless of the value of ϕ. The fact that ϕ must take on a particular value at any given time and place tells us that the $U(1)_{PQ}$ symmetry is spontaneously broken.

To resolve the strong CP problem, some quarks (either those in the Standard Model, or some other heavy exotic particles that carry QCD color) must be charged under the $U(1)_{PQ}$ symmetry. Through a loop of these quarks, the φ field couples to a pair of gluons in a way that involves what is known as a color anomaly. Anomalies arise when a quantum field theory does not respect a symmetry that is manifest in the classical theory. In this case, the classical symmetry is violated by the production of a gluon pair. As a result of this anomaly, the degree of CP violation that is manifest among the strong interactions can be directly related to the phase of the scalar field, $\theta_{QCD} = \phi \times (N/f_{PQ})$, where N is a model-dependent factor that we will return to below. The key point here is that instead of being a constant, θ_{QCD} is now a quantity that can evolve and change—it is a dynamical variable.

At high temperatures, the value of ϕ, and therefore of θ_{QCD}, is free to roll around the bottom of the potential. This, however, does not solve our problem, as there is no particular reason for it to roll toward or remain at the measured value of $\theta_{QCD} \approx 0$. At temperatures below $T \sim \Lambda_{QCD}$, however, non-perturbative effects cause the shape of this potential to become slightly tilted to one side, breaking the degeneracy that had previously existed between all values of ϕ. More specifically, the potential found throughout the high-temperature minima, $\varphi = (f_{PQ}/\sqrt{2})\, e^{i\phi/f_{PQ}}$, is now a function of the complex phase, and is given by the following expression [410]:

$$V(\phi) \approx V_0 - m_\pi^2 f_\pi^2 \left[1 - \frac{4 m_u m_d}{(m_u + m_d)^2} \sin^2(N\phi/2f_{PQ})\right]^{1/2}, \tag{17.42}$$

where $V_0 = m_\pi^2 f_\pi^2$ is a constant chosen such that $V = 0$ at the minimum, and $f_\pi \approx 92$ MeV is a quantity known as the pion decay constant. As a result of this tilting, the potential is now minimized at $\phi = 0$, providing us with a dynamical explanation for why $\theta_{QCD} = 0$, and thus for why the neutron's electric dipole moment is so much smaller than we would have otherwise expected it to have been.

Shortly after Peccei and Quinn introduced their solution to the strong CP problem, Steven Weinberg [411] and Frank Wilczek [412] each independently pointed out that the broken $U(1)_{PQ}$ symmetry at the heart of this scenario implies the existence of a new particle. This comes from what is known as Goldstone's theorem, which states that any spontaneously broken continuous symmetry corresponds to a spin-zero massless particle—a Nambu-Goldstone boson [413, 414]. In the case of the broken $U(1)_{PQ}$ symmetry, the Nambu-Goldstone boson is a pseudoscalar called the axion.

The tilting of the potential that takes place at $T \sim \Lambda_{QCD}$ causes the axion to acquire a small mass, on the order of $m_a \sim \Lambda_{QCD}^2/f_{PQ}$. We can calculate the value of this mass from the shape of the potential given above [410]:

$$m_a^2 = \frac{\partial^2 V}{\partial \phi^2}(\phi = 0) = \frac{m_u m_d}{(m_u + m_d)^2} \frac{m_\pi^2 f_\pi^2}{(f_{PQ}/N)^2} \tag{17.43}$$

$$\longrightarrow m_a = \frac{(m_u m_d)^{1/2}}{(m_u + m_d)} \frac{m_\pi f_\pi}{(f_{\rm PQ}/N)}$$
$$\approx 5.7\,{\rm keV} \times \left(\frac{1\,{\rm TeV}}{f_{\rm PQ}/N}\right),$$

where in the last step, I adopted $m_u/m_d \approx 0.49$ for the ratio of the light quark masses.

The couplings of the axion to Standard Model particles are proportional to $f_{\rm PQ}^{-1}$. Of particular importance is the coupling of the axion to two photons, which arises through the mixing of the axion with the π^0 and η mesons. This two-photon coupling can be written in terms of electric and magnetic fields as follows:

$$\mathcal{L}_{a\gamma\gamma} = g_{a\gamma\gamma}\, a\, \vec{E} \cdot \vec{B}, \tag{17.44}$$

where the dimensionful coupling constant is given by

$$g_{a\gamma\gamma} = \frac{\alpha g_\gamma}{\pi\,(f_{\rm PQ}/N)} \tag{17.45}$$
$$\approx 2.3 \times 10^{-6}\,{\rm GeV}^{-1} \left(\frac{g_\gamma}{1}\right) \left(\frac{1\,{\rm TeV}}{f_{\rm PQ}/N}\right).$$

The quantity g_γ in this expression is an order 1 number whose value depends on the detailed particle content of the theory. The axion's coupling to two photons is important because, among other reasons, it provides a way to potentially detect axions, through their conversion into photons in the presence of an external magnetic field [415] (see Sec. 18.4).

As originally envisioned, the scale of Peccei-Quinn symmetry breaking was expected to be near the weak scale, $f_{\rm PQ} \sim v$, leading to an axion with a mass on the order of $m_a \sim \mathcal{O}(10\,{\rm keV})$, and with couplings large enough to be accessible at a variety of laboratory experiments. In particular, keV-scale axions would have been produced in observable quantities at nuclear reactors, and would have been readily detectable in beam-dump experiments. In response to the constraints that have been derived from these experiments, the value of $f_{\rm PQ}$ is now generally taken to be much larger than the electroweak scale.

The properties of axions are also subject to a variety of astrophysical constraints. In particular, large numbers of axions would be produced in the interiors of stars if $f_{\rm PQ}$ is not very large [416]. Unlike photons, any axions that are created within a stellar core will readily escape, carrying away energy and cooling the star more rapidly than would have otherwise been predicted. Observations of horizontal branch red giant stars have been used to constrain $g_{a\gamma\gamma} \lesssim 6.6 \times 10^{-11}\,{\rm GeV}^{-1}$ [417], corresponding to $f_{\rm PQ}/N \gtrsim 3.5 \times 10^7\, g_\gamma$ GeV. Furthermore, the number of neutrino events observed in conjunction with supernova 1987A (25 events over ~ 10 s) is in good agreement with the predictions of core-collapse models, limiting how many axions could have been produced through nucleon-nucleon scattering, $N + N \to N + N + a$. This observation would appear to require $f_{\rm PQ}/N \gtrsim 10^8$ GeV and (applying Eq. 17.43) $m_a \lesssim$ 0.1–0.01 eV [418]. These constraints, along with cosmological considerations that we will discuss below, motivate scenarios in which the Peccei-Quinn scale is in the vicinity of $f_{\rm PQ} \sim 10^9$–10^{13} GeV, leading to a so-called "invisible" axion that is both very feebly interacting, and very light, $m_a \approx 5.7\,\mu{\rm eV} \times [10^{12}\,{\rm GeV}/(f_{\rm PQ}/N)]$.

While many different axion models have been proposed, two particular benchmarks stand out. In the first of these, known as the KSVZ (Kim-Shifman-Vainshtein-Zahkarov)

model [419, 420], there exist heavy quark-like particles with QCD color, but without any weak couplings or electric charge. The couplings of the axion to these exotic quarks induce interactions between the axion and various Standard Model particles. The other widely studied axion model, called the DFSZ (Dine-Fisher-Srednicki-Zhitnisky) model [421, 422], incorporates two Higgs doublets and an additional scalar that acquires a vacuum expectation value near the scale of Peccei-Quinn symmetry breaking. The axion then couples to the Standard Model through this scalar, which interacts with the Standard Model fermions through its couplings to the Higgs doublets. In the context of a specific model, such as either of these benchmarks, we can calculate the value of g_γ that, in turn, goes into determining the axion's coupling to two photons:

$$g_\gamma = \frac{1}{2}\left(\frac{E}{N} - \frac{2\,[4 + (m_u/m_d)]}{3\,[1 + (m_u/m_d)]}\right). \tag{17.46}$$

The quantities N and E are model-dependent integers known as the axion color anomaly and the axion electromagnetic anomaly. In the KSVZ model, $E = 0$, leading to $g_\gamma \approx -0.97$, while in the case of DFSZ, $E/N = 8/3$, and thus $g_\gamma \approx 0.36$. Many other axion models have been proposed, featuring a range of values for E/N [423, 424, 425].

Although the axion is not entirely stable, the lifetime of this particle can be very long. In particular, the coupling given in Eq. 17.45 allows the axion to decay into a pair of photons with a lifetime given by:

$$\begin{aligned}
\tau_{a\to\gamma\gamma} &= \frac{64\pi}{m_a^3 g_{a\gamma\gamma}^2} \qquad (17.47)\\
&= \frac{64\pi^3 (f_{\rm PQ}/N)^2}{m_a^3 \alpha^2 g_\gamma^2}\\
&\approx \begin{cases} 4\times 10^{42}\ {\rm yr} \times \left(\dfrac{f_{\rm PQ}/N}{10^{12}\ {\rm GeV}}\right)^2 \left(\dfrac{5.7\ \mu eV}{m_a}\right)^3 & \text{(KSVZ)},\\[2ex] 3\times 10^{43}\ {\rm yr} \times \left(\dfrac{f_{\rm PQ}/N}{10^{12}\ {\rm GeV}}\right)^2 \left(\dfrac{5.7\ \mu eV}{m_a}\right)^3 & \text{(DFSZ)}. \end{cases}
\end{aligned}$$

Using Eq. 17.43 to relate $f_{\rm PQ}$ to m_a, we can express the axion's lifetime in terms of either of these quantities, finding that $\tau_a \propto f_{\rm PQ}^5$ and $\tau_a \propto m_a^{-5}$. For axions lighter than $m_a \sim \mathcal{O}(10\ {\rm eV})$, the lifetime of these particles exceeds the age of the universe. In light of the axion's very feeble couplings and long lifetime, it's easy to see why this particle has so often been studied as a possible candidate for the dark matter of our universe.

Before moving on to consider axions as dark matter, I want to briefly differentiate the "axions" that we have been discussing in this section from the broader class of "axion-like particles," or ALPs as they are often known. The term axion—or QCD axion—refers specifically to the particles that we have been talking about here, which arise within the context of the Peccei-Quinn solution to the strong CP problem. In contrast, many papers have been written about a broader class of light pseudoscalars with similar properties to those of the QCD axion. One central distinction is that while the mass of the QCD axion is directly related to the scale of Peccei-Quinn symmetry breaking, and thus to the strength of its couplings (see Eq. 17.43), these two quantities are often treated as independent in discussions of axion-like particles. You may have noticed that I have written some of the

equations in this section in terms of both m_a and $f_{\rm PQ}$, and perhaps wondered why I did not simplify these expressions by applying Eq. 17.43. The reason is that, by presenting them as written, these equations can be applied not only to QCD axions, but also to a broader range of axion-like particles.

17.3.1 Axion production in the early universe

Thermal Production Based on the discussion of dark matter production in Chapter 9, I will begin this section by considering the possibility that a substantial population of axions may have been produced thermally in the early universe. The most important process for thermal axion production is $N + \pi \to N + a$ [426], whose thermally averaged cross section is roughly given by

$$\langle \sigma\, v \rangle \sim \frac{T^2}{(f_{\rm PQ}/N)^2 m_\pi^2}. \tag{17.48}$$

Combining this with the equilibrium abundance of nucleons, we arrive at the following rate for the production of axions:

$$\begin{aligned} \Gamma_{N+\pi\to N+a} &= \langle \sigma\, v \rangle\, n_N^{\rm Eq} \\ &\sim \frac{T^2}{(f_{\rm PQ}/N)^2 m_\pi^2}\, \frac{1}{\sqrt{2}} \left(\frac{m_N T}{\pi}\right)^{3/2} e^{-m_N/T}. \end{aligned} \tag{17.49}$$

Comparing this to the rate of Hubble expansion, we obtain

$$\begin{aligned} \frac{\Gamma_{N+\pi\to N+a}}{H} &\sim \frac{3\sqrt{5} m_N^{3/2} T^{3/2} M_{\rm Pl}\, e^{-m_N/T}}{2\sqrt{2}\pi^3 (f_{\rm PQ}/N)^2 m_\pi^2 g_\star(T)^{1/2}} \\ &\sim 4 \times 10^{-8} \times \left(\frac{10^{12}\,{\rm GeV}}{f_{\rm PQ}/N}\right)^2, \qquad \text{(evaluated at } T \sim \Lambda_{\rm QCD}\text{)}, \end{aligned} \tag{17.50}$$

where in the last line I have evaluated this ratio at $T \sim \Lambda_{QCD}$, where it is largest. Unless $f_{\rm PQ}/N \lesssim 2 \times 10^8$ GeV (corresponding to $m_a \gtrsim 0.03$ eV), the axion production rate will remain below the Hubble rate throughout the early universe. In the case of a heavier and less feebly interaction axion, $m_a \gtrsim 0.03$ eV, the production rate can exceed the Hubble rate, allowing the abundance of these particles to reach their equilibrium value. Just as we did in the case of the cosmic neutrino background (see Sec. 4.5), we can determine the current temperature and density of these thermal axions by applying the principle of entropy conservation:

$$T_a = T \left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})}\right)^{1/3}, \tag{17.51}$$

where $T_{\rm dc}$ is the temperature at which the axions decoupled from the thermal bath. This temperature corresponds to a density that is given by

$$\rho_a = m_a \frac{\zeta(3) T_a^3}{\pi^2} \tag{17.52}$$

$$= m_a \frac{\zeta(3) T_{\rm CMB}^3}{\pi^2} \left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_{\rm dc})} \right)$$
$$\approx 8 \times 10^{-8}\,{\rm GeV/cm^3} \times \left(\frac{m_a}{\rm eV} \right) \left(\frac{10}{g_{\star,S}(T_{\rm dc})} \right).$$

From this, we learn that the density of thermal axions will be well below the dark matter density unless $m_a \gtrsim 20\,{\rm eV}$.

There are multiple reasons for why thermal axions are not a tenable dark matter candidate. First, axions of this mass would be unstable on cosmological timescales (see Eq. 17.47). Furthermore, a thermal population of particles this light would not contribute to the abundance of cold dark matter, but would instead be quite hot, in conflict with large-scale structure. Finally, axions in this mass range are ruled out by a variety of constraints related to stellar evolution. Thus, if axions are to make up the dark matter of our universe, we will need to rely on a nonthermal mechanism for their production.

Before moving on to discuss nonthermal production mechanisms for axions, let's consider the impact of thermal axions on the expansion history of our universe. As a starting point, we can consider the case in which the thermal axions are light enough to have been relativistic at the time of photon decoupling, $m_a \lesssim 0.3\,{\rm eV} \times [3.9/g_{\star,S}(T_{\rm dc})]^{1/3}$, but heavy enough to have reached equilibrium around the time of the QCD phase transition, $m_a \gtrsim 0.03\,{\rm eV}$. In this mass range, the energy density in thermal axions can be written as

$$\rho_a \approx \frac{\pi^2 T_a^4}{30} = \frac{\pi^2\, T^4}{30} \left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})} \right)^{4/3}. \tag{17.53}$$

Using Eq. 7.27, we can express this energy density in terms of a contribution to $N_{\rm eff}$:

$$\Delta N_{\rm eff} = \frac{8}{7} \left(\frac{11}{4} \right)^{4/3} \frac{\rho_a}{\rho_\gamma} = \frac{4}{7} \left(\frac{11}{4} \right)^{4/3} \left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_{\rm dc})} \right)^{4/3} \approx 0.15 \times \left(\frac{30}{g_{\star,S}(T_{\rm dc})} \right)^{4/3}. \tag{17.54}$$

Axions that decoupled from the thermal bath after the QCD phase transition would be expected to provide a relatively significant contribution to the energy density of the early universe, $\Delta N_{\rm eff} \sim 0.1$–$0.2$. Although this range is consistent with current measurements of the CMB and baryon acoustic oscillations, $\Delta N_{\rm eff} = -0.056^{+0.34}_{-0.32}$, it is also within the projected sensitivity of future observations [427]. Somewhat lighter axions, which do not reach equilibrium in the early universe, could still be produced through freeze in, leading to a smaller contribution to $N_{\rm eff}$.

Axions from Vacuum Misalignment In addition to any thermal population that may have been generated in the early universe, very light axions could also have been efficiently produced through what is known as the misalignment mechanism [428, 429, 430]. The

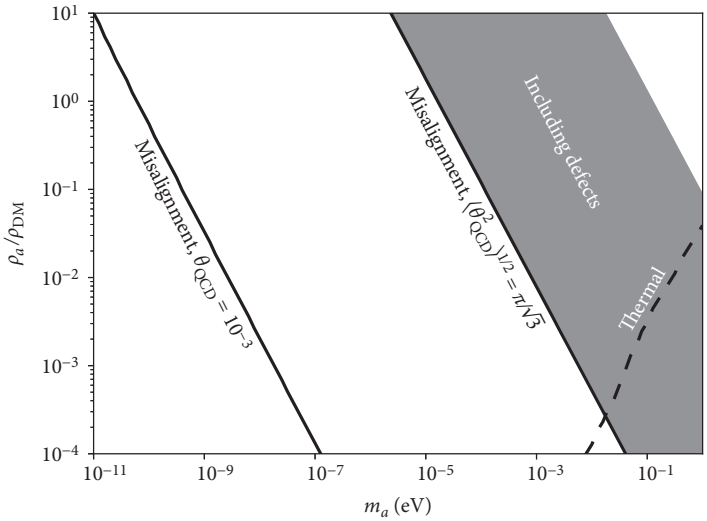


Figure 17.6 The abundances of axions that result from thermal production, vacuum misalignment, and the decays of topological defects. Axions more massive than 0.06 eV (KSVZ) or 0.046 eV (DFSZ) are excluded by the neutrino flux observed from supernova 1987A [418].

idea here is that the slight tilt in the axion potential initially causes ϕ to evolve toward zero, where the potential is minimized. The value of ϕ, however, inevitably overshoots this value, and ends up oscillating around the minimum. As the value of ϕ relaxes toward this minimum, the energy that had been stored in the initial configuration of the vacuum becomes transferred into the production of nonthermal axions.

Near the minimum of Eq. 17.42 (i.e., near $|\phi| \sim 0$), the potential is approximately quadratic, $V(\phi) \approx (m_a^2/2)\,\phi^2 = (m_a^2/2)(f_{\rm PQ}/N)^2\,\theta_{\rm QCD}^2$. Including the mass of the axion, the equation of motion in this case is that of a damped harmonic oscillator:

$$\ddot{\phi} + 3H\dot{\phi} + m_a^2(T)\phi = 0. \tag{17.55}$$

Note that, unlike in the case of a standard damped harmonic oscillator, the mass of the axion that appears in the third term of this expression is a function of temperature, and thus evolves. Throughout the remainder of this section, I will explicitly write the temperature dependance of the axion's mass, $m_a(T)$, differentiating this quantity from the axion's low-temperature mass, $m_a \equiv m_a(T=0)$.

At temperatures well above the QCD phase transition, the axion's mass is approximately given by [431]

$$m_a(T) \approx 0.1\, m_a \left(\frac{\Lambda_{\rm QCD}}{T}\right)^{3.7}, \qquad T \gg \Lambda_{\rm QCD}, \tag{17.56}$$

From this expression, we see that at high temperatures, $T \gg \Lambda_{\rm QCD}$, the axion is nearly massless, and thus the solutions to the equation of motion feature oscillations that are damped by a factor of $e^{-3Ht/2}$. This quickly puts an end to any evolution of ϕ, forcing it to freeze in place and take on a nearly constant value. As the temperature drops, however, ϕ becomes dislodged and begins to oscillate around the minimum. In particular, once $m_a(T) > 3H/2$, the solutions to Eq. 17.55 transition from those associated with an overdamped harmonic oscillator to those of an underdamped harmonic oscillator, featuring an angular frequency given by $\omega = [m_a^2(T) - (3H/2)^2]^{1/2}$.

The energy density of the axion field as it oscillates around the minimum of its potential is given by

$$\rho_a = \frac{\dot{\phi}^2}{2} + \frac{m_a^2(T)\,\phi^2}{2}. \tag{17.57}$$

From the virial theorm (see Sec. 9.1), it follows that

$$\langle \dot{\phi}^2 \rangle = m_a^2(T) \langle \phi^2 \rangle = \rho_a, \tag{17.58}$$

where the brackets represent time averages over one or more oscillations. Multiplying Eq. 17.55 by $\dot{\phi}$, and replacing $\dot{\phi}^2$ with ρ_a, we have

$$\ddot{\phi}\,\dot{\phi} + 3H\rho_a + m_a^2(T)\,\phi\,\dot{\phi} = 0. \tag{17.59}$$

From Eq. 17.57, it follows that

$$\begin{aligned} \ddot{\phi}\,\dot{\phi} + m_a^2(T)\,\phi\,\dot{\phi} &= \dot{\rho}_a - m_a(T)\dot{m}_a\phi^2 \\ &= \dot{\rho}_a - \frac{\dot{m}_a(T)}{m_a(T)}\,\rho_a. \end{aligned} \tag{17.60}$$

Combining this with Eq. 17.59, we arrive at

$$\dot{\rho}_a - \frac{\dot{m}_a(T)}{m_a(T)}\rho_a + 3H\rho_a = 0, \tag{17.61}$$

whose solution is $\rho_a \propto m_a(T)/a^3$, where a is the scale factor.

This important result demonstrates that, up to time or temperature variations in $m_a(T)$, the comoving energy density of the axion field remains constant as the universe expands. In other words, once the temperature of the universe has cooled enough for the axion mass to take on a nearly constant value ($T \ll \Lambda_{\rm QCD}$), the energy density of the axion field evolves just like nonrelativistic matter. From this point onward, it is useful to think about the axion field as a gas of nonrelativistic particles, with a number density given by $n_a = \rho_a/m_a$. It is also worth emphasizing that, despite being very light, any axions that are produced through the misalignment mechanism will be very cold, leading to acceptable large-scale structure.

The energy density of these axions is directly related to the energy that was initially stored in the misalignment of the axion field relative to its minimum at $\phi = 0$. To determine the final abundance of these particles, we must first establish when the axion field started to oscillate. To accomplish this, we set the temperature-dependent mass of the axion (see Eq. 17.56) equal to $3H/2$ and solve for temperature. From this, it follows that the axion field begins to oscillate approximately when

$$\begin{aligned} 0.1\,m_a \left(\frac{\Lambda_{\rm QCD}}{T_i}\right)^{3.7} &\sim \frac{3}{2}\left(\frac{8\pi^3 g_\star(T_i)\,T_i^4}{90\,M_{\rm Pl}^2}\right)^{1/2} \\ \longrightarrow\; T_i &\sim \left(\frac{\sqrt{5}\,m_a\,\Lambda_{\rm QCD}^{3.7}\,M_{\rm Pl}}{10\,\pi^{3/2}\,g_\star(T_i)^{1/2}}\right)^{1/5.7} \\ &\approx 1.0\,{\rm GeV} \times \left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{0.175} \left(\frac{70}{g_\star(T_i)}\right)^{0.088}. \end{aligned} \tag{17.62}$$

At the time that the axion field was just starting to oscillate, $m_a(T) \sim 3H/2$, the energy density in axions was approximately given by

$$\begin{aligned}\rho_a(T_i) &\sim V \sim \frac{m_a^2(T_i)\,\phi^2}{2} \\ &= \frac{m_a^2(T_i)\,(f_{\rm PQ}/N)^2\,\theta_{\rm QCD}^2}{2} \\ &\approx 4.4\times 10^{-11}\,{\rm GeV}^4 \times \left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{0.70} \left(\frac{f_{\rm PQ}/N}{5.7\times 10^{12}\,{\rm GeV}}\right)^2 \left(\frac{g_\star(T_i)}{70}\right)^{0.65} \theta_{\rm QCD}^2,\end{aligned} \tag{17.63}$$

and thus

$$\begin{aligned}n_a(T_i) &= \frac{\rho_a(T_i)}{m_a(T_i)} \\ &\sim 2.7\times 10^{7}\,{\rm GeV}^3 \times \left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{0.35} \left(\frac{f_{\rm PQ}/N}{5.7\times 10^{12}\,{\rm GeV}}\right)^2 \left(\frac{g_\star(T_i)}{70}\right)^{0.32} \theta_{\rm QCD}^2.\end{aligned} \tag{17.64}$$

Keep in mind that $\theta_{\rm QCD}$ in these expressions represents the value of this angle before the axion field began to evolve toward the new ground state at $\theta_{\rm QCD} = 0$. In other words, $\theta_{\rm QCD}$ is the angle by which the vacuum was initially misaligned relative to the ground state.

From entropy conservation, it follows that the scale factor has increased by a factor of $(T_i/T_{\rm CMB})\,[g_{\star,S}(T_i)/g_{\star,S}(T_{\rm CMB})]^{1/3}$ since the time at which the temperature of the universe was equal to T_i. Thus the density of these axions in the universe today is approximately given by

$$\begin{aligned}\rho_a &= n_a\,m_a \\ &= n_a(T_i)\,m_a \left(\frac{T_{\rm CMB}}{T_i}\right)^3 \left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_i)}\right) \\ &\sim 2.5\times 10^{-6}\,{\rm GeV/cm}^3 \times \left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{0.825} \left(\frac{f_{\rm PQ}/N}{5.7\times 10^{12}\,{\rm GeV}}\right)^2 \left(\frac{g_\star(T_i)}{70}\right)^{-0.42} \theta_{\rm QCD}^2.\end{aligned} \tag{17.65}$$

Relating m_a to $f_{\rm PQ}/N$ according to Eq. 17.43, we arrive at

$$\rho_a \sim 2.5\times 10^{-6}\,{\rm GeV/cm}^3 \times \left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{-1.175} \left(\frac{g_\star(T_i)}{70}\right)^{-0.42} \theta_{\rm QCD}^2. \tag{17.66}$$

In the course of this calculation, we have made a number of approximations and estimates. A more detailed version of this calculation yields the following result for the abundance of axions from the misalignment of the vaccum [432, 433, 434, 435]:

$$\rho_a \approx 1.1\times 10^{-5}\,{\rm GeV/cm}^3 \times \left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{-1.175} \left(\frac{g_\star(T_i)}{70}\right)^{-0.42} F\,\theta_{\rm QCD}^2, \tag{17.67}$$

where F is an order 1 factor that accounts for aharmonics in the axion potential, and is a function of both $\theta_{\rm QCD}$ and $f_{\rm PQ}/N$.

Unlike most other dark matter candidates, axions are produced in an abundance that depends on an *environmental parameter*, $\theta_{\rm QCD}$. By environmental, I mean that this parameter can take on different values at different locations throughout space. With this in mind, we can consider two rather different scenarios. First, if the Peccei-Quinn symmetry is broken after inflation, then our universe will contain a very large number of regions, each with a different initial value of $\theta_{\rm QCD}$. These regions will be necessarily small because they are limited to the size of the horizon at the time of Peccei-Quinn breaking. Since there is no mechanism for causally disconnected regions to have correlated values of their misalignment angle, macroscopic regions in the universe today will exhibit a density of axions that can be calculated using the value of $\theta_{\rm QCD}^2$ averaged over a large number of such regions,

$$\begin{aligned}\langle\theta_{\rm QCD}^2\rangle &\equiv \frac{1}{2\pi}\int_{-\pi}^{\pi} d\theta_{\rm QCD}\,\theta_{\rm QCD}^2 \\ &= \frac{\pi^2}{3},\end{aligned} \tag{17.68}$$

leading to the following density of axions produced through vacuum misalignment:

$$\rho_a \approx 3.5\times 10^{-5}\,{\rm GeV/cm^3}\times\left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{-1.175}\left(\frac{g_\star(T_i)}{70}\right)^{-0.42} F. \tag{17.69}$$

Thus, for $F\approx 1$, one obtains an axion abundance that is equal to the measured dark matter density ($\rho_{\rm DM}\approx 1.3\times 10^{-6}\,{\rm GeV/cm^3}$) for a mass of approximately $m_a\sim 20\,\mu{\rm eV}$.

Alternatively, it is plausible that Peccei-Quinn symmetry breaking could have occurred prior to inflation. In this case, the process of inflation would have taken small regions of space (each containing an approximately constant value of $\theta_{\rm QCD}$) and stretched them into volumes that are larger than the size of the observable universe today. In such a scenario, $\theta_{\rm QCD}$ is equally likely to take on any value between π and $-\pi$, and thus we can only make statistical statements about the abundance of axions that was produced through vacuum misalignment in our observable universe. While we expect most regions of space in this scenario to feature an initial value of $\theta_{\rm QCD}$ that is similar to the average value, there will exist some regions where this quantity is much smaller than the average value. In some theoretical frameworks, one expects $f_{\rm PQ}$ to be near the GUT or Planck scale, and thus for the axions to be very light. In this case, most regions of space (those with $\theta_{\rm QCD}\sim\mathcal{O}(1)$) will contain a very large density of axions. In those typical regions, where the axion density is much greater than the dark matter density found in our universe, the process of star formation will be highly suppressed, limiting the probability that life will emerge [436]. This potentially provides us with an anthropic explanation for why $\theta_{\rm QCD}$ might be very small in our observable universe, or in any other region of space in which life is found to exist [436].

In scenarios where the Peccei-Quinn symmetry was broken prior to inflation, one expects quantum fluctuations in the axion field to lead to variations in the axion density that do not necessarily trace the densities of radiation or other forms of energy [437, 438]. This leads to isocurvature perturbations (see Sec. 10.8), the presence of which is strongly constrained by the measured temperature anisotropies in the CMB. Quantitatively, in a scenario with pre-inflationary Peccei-Quinn breaking, and in which axions make up all of the dark mater, the lack of observed isocurvature in the CMB allows us to place the following constraint on the Hubble rate during inflation [439, 440, 441]:

$$H_I \lesssim 6.2\times 10^7\,{\rm GeV}\times\left(\frac{10^{-6}\,{\rm eV}}{m_a}\right)^{0.42}. \tag{17.70}$$

From the first Friedmann equation, we can relate this to a constraint on the energy density during inflation, which in turn can be expressed in terms of an effective temperature, sometimes called the energy scale of inflation:

$$T_I \equiv \left(\frac{30\rho_I}{\pi^2 g_\star(T_I)}\right)^{1/4} \lesssim 2 \times 10^{13}\,\text{GeV} \times \left(\frac{10^{-6}\,\text{eV}}{m_a}\right)^{0.21} \left(\frac{106.75}{g_\star(T_I)}\right). \tag{17.71}$$

From this, we conclude that pre-inflationary Peccei-Quinn breaking is possible only if the energy density during inflation, as well the subsequent temperature of reheating, was not too high.

Axions from Topological Defects In addition to any axions that were produced thermally or through vacuum misalignment, an abundance of cold axions could also have been produced through the decays of topological defects, such as strings or domain walls [442, 443, 444, 445]. Compared to other axion populations, the abundance of these axions is more model-dependent, and is subject to significant uncertainties [434, 446, 447, 448, 449, 450, 451, 452]. Depending on the details that are assumed, the decays of topological defects could have played the dominant role in the production of axions in the early universe [453, 454, 455], or may have been largely irrelevant [456, 457].

As discussed in Sec. 12.7, topological strings arise in theories in which a $U(1)$ gauge symmetry is spontaneously broken. The same is true in the case of a global $U(1)$ symmetry, and the $U(1)_{\rm PQ}$ symmetry associated with the axion is no exception. Topological defects arise within this class of models because the initial values of $\theta_{\rm QCD}$ that are mapped across a given volume of space can feature nontrivial topologies. As a consequence, a network of axion strings and domain walls develops as the Peccei-Quinn symmetry is broken. If this takes place prior to inflation, the resulting topological defects will be diluted away and thus play no significant role in the production of axions. In contrast, if the Peccei-Quinn symmetry is broken after inflation, the decays of the strings and domain walls will lead to the production of a sizable axion abundance.

The energy density in long-lived domain walls evolves as $\rho_{\rm DW} \propto a^{-2}$. As a result, such objects could easily come to dominate the energy density of the early universe, leading to an unacceptably large abundance of axions. With this in mind, we will focus here on scenarios in which the domain walls decay quickly, leaving axion strings to be the main source of axion production.

Whereas other kinds of topological strings can dissipate their energy into gravitational waves, axion strings shrink, straighten, combine, and otherwise lose their energy by radiating axions. As a network of axion strings evolves, it will tend toward a state in which the total energy density in these objects is given by $\rho_S \sim \mu/t^2$, where $\mu \sim (f_{\rm PQ}/N)^2 \ln(d f_{\rm PQ}/N)$ is a given string's energy per unit length, d is the typical distance between strings, and t is the time. Notice that since $\rho \propto T^4 \propto t^{-2}$ during radiation domination, the string network constitutes a roughly constant fraction of the total energy density throughout the early universe.

A precise calculation of the axion abundance produced through string decays would be quite technical, and is beyond the scope of the discussion at hand. Instead, I will provide a back-of-the-envelope estimate here, which will yield approximately the correct result. Roughly speaking, the total energy density that is stored in the axion string network will be radiated as axions approximately once per Hubble time. With this in mind, we can write the following expression for the change in the number density of axions per unit entropy

density:

$$\int \frac{dn_a}{s} \sim \int \left(\frac{\mu}{t^2\,\omega}\right)\left(\frac{45}{2\pi^2 g_{\star,S}(T) T^3}\right)\frac{dt}{t_H}, \tag{17.72}$$

where ω is the average energy of a radiated axion, and $t_H \sim H^{-1}$. The first term in brackets is the total number of axions radiated by the string network in a Hubble time, while the second term is 1 over the entropy density of the thermal bath. Writing $t = H^{-1} = [(3M_{\rm Pl}^2/8\pi^3 g_\star(T) T^4]^{1/2}$ and thus $dt \approx -[3M_{\rm Pl}^2/8\pi^3 g_\star(T)]^{1/2}\, 2\, dT/T^3$, we can express the above equation as follows:

$$\begin{aligned}\frac{n_a}{s} &\sim \frac{120\pi}{M_{\rm Pl}^2}\int_{T_i}^{f_{\rm PQ}/N}\frac{\mu}{\omega}\,\frac{g_\star(T)}{g_{\star,S}(T)}\,dT \\ &\sim \frac{60\sqrt{3}}{M_{\rm Pl}\sqrt{2\pi}}\int_{T_i}^{f_{\rm PQ}/N}\mu\,\frac{g_\star(T)^{1/2}}{g_{\star,S}(T)}\,\frac{dT}{T^2} \\ &\sim \frac{60\sqrt{3}\ln(d f_{\rm PQ}/N)}{\sqrt{2\pi}\,g_\star(T_i)^{1/2}}\,\frac{(f_{\rm PQ}/N)^2}{M_{\rm Pl}\,T_i},\end{aligned} \tag{17.73}$$

where T_i is again the temperature at which $m_a(T) \sim 3H/2$ and, in the second line, I have adopted an estimate of $\omega \sim t^{-1}$. Combining this with the expression for T_i given in Eq. 17.62, we can write the current density of axions from string decays as follows:

$$\begin{aligned}\rho_a &= n_a\, m_a \\ &= n_a(T_i)\left(\frac{T_{\rm CMB}}{T_i}\right)^3\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_i)}\right)m_a \\ &\sim \frac{n_a(T_i)}{s(T_i)}\,s(T_i)\left(\frac{T_{\rm CMB}}{T_i}\right)^3\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_i)}\right)m_a \\ &\sim \frac{60\sqrt{3}\ln(d f_{\rm PQ}/N)}{\sqrt{2\pi}\,g_\star(T_i)^{1/2}}\,\frac{(f_{\rm PQ}/N)^2}{M_{\rm Pl}}\left(\frac{10\,\pi^{3/2}\,g_\star(T_i)^{1/2}}{\sqrt{5}\,m_a\,\Lambda_{\rm QCD}^{3.7}\,M_{\rm Pl}}\right)^{1/5.7}\left(\frac{2\pi^3 g_{\star,S}(T_i)\,T_i^3}{45}\right) \\ &\quad\times\left(\frac{T_{\rm CMB}}{T_i}\right)^3\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_i)}\right)m_a \\ &\approx 1.2\times10^{-4}\,{\rm GeV/cm^3}\times\left(\frac{f_{\rm PQ}/N}{5.7\times10^{12}\,{\rm GeV}}\right)^2\left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{0.825}\left(\frac{g_\star(T_i)}{70}\right)^{0.59}\ln(d f_{\rm PQ}/N) \\ &\approx 1.2\times10^{-4}\,{\rm GeV/cm^3}\times\left(\frac{m_a}{10^{-6}\,{\rm eV}}\right)^{-1.175}\left(\frac{g_\star(T_i)}{70}\right)^{0.59}\ln(d f_{\rm PQ}/N),\end{aligned} \tag{17.74}$$

where in the last step I have applied Eq. 17.43 to relate $f_{\rm PQ}/N$ to m_a. Notice that the abundance of axions that results from string decays depends on the axion mass in the same way that the abundance from vacuum misalignment does, $\rho_a \propto m_a^{-1.175}$ (see Eq. 17.67). When comparing this expression to the corresponding contribution from vacuum misalignment, we find that the decays of topological defects are likely to be the

dominant mechanism for axion production in the early universe, at least if the Peccei-Quinn symmetry is broken after inflation. More specifically, this calculation yields an abundance of axions that is equal to observed dark matter density for $m_a \sim 50\ \mu$eV.

While the estimate described above yields a result that is similar to those found in more detailed numerical calculations (see, for example, Ref. [458]), such calculations can also be sensitive to the details of the model in question, and continue to suffer from numerical uncertainties. For axion models with $N = 1$ (such as the KSVZ model), the domain walls decay quite quickly, leaving strings to play the dominant role and leading to an abundance of axions that is equal to the dark matter density for $m_a \sim$20–4,000 μeV. For models with $N > 1$ (such as the DFSZ model), the domain walls are typically rather long-lived, leading to the production of a very large axion abundance. If no mechanism to remove these domain walls exists [459], the axion abundance in this case will equal the measured dark matter density for masses in the range of $m_a \sim$0.6–130 meV.

17.4 Dark matter in hidden sectors

For decades, WIMPs were the most widely studied class of dark matter candidates. As the constraints on such particles from both colliders and direct detection experiments (see Chapter 18) have become increasingly stringent, however, alternative scenarios have become steadily more attractive. For example, the dark matter could be part of a "hidden sector," which consists of a number of particle species that do not carry any Standard Model gauge charges. If the dark matter particles annihilate into lighter hidden sector states that then decay into Standard Model particles, an acceptable relic abundance could easily be obtained, while remaining consistent with all existing constraints.

Hidden sector particles could decay into, and otherwise interact with, Standard Model particles through what are known as "portal couplings." Some minimal examples of such couplings include the Higgs portal, vector portal, and lepton portal, which result from the mixing of a hidden sector state with the Standard Model Higgs boson, hypercharge gauge boson, or right-handed neutrino [460]. Depending on the size of the effective portal coupling, scattering processes could keep the hidden sector particles in kinetic equilibrium with the Standard Model bath, or leave these two groups of particles completely decoupled from one another [461, 462, 463, 464].

As an example, consider a scenario in which the dark matter candidate, X, annihilates into a pair of other hidden sector states, $XX \to YY$, followed by the Y particles decaying through a portal coupling into some combination of Standard Model particles. If the portal coupling is not too small, kinetic equilibrium will be maintained between the sectors, and the decays of the Y particles will occur promptly, $\Gamma_Y \gtrsim H$. In this case, the relic abundance of the X population will be similar to that of an ordinary WIMP undergoing thermal freeze out, as described in Sec. 9.5.2. In contrast, if the portal coupling is very small, then the two sectors will be decoupled from one another, each evolving with its own temperature. Furthermore, if the portal coupling is small enough, the Y particles will be long-lived, potentially leading to an era of early matter-domination, followed by the late-time reheating of the Standard Model bath.

Additional reading

When it comes to supersymmetry, I am particularly fond of Fox's lectures from the TASI summer school [465], as well as Martin's classic Supersymmetry Primer [372]. For an even

deeper dive into *Weak Scale Supersymmetry*, the textbook by Baer and Tata is a good reference. For sterile neutrinos as a dark matter candidate, nice reviews have been written by Abazajian [466] and Kusenko [467]. For axions, I recommend the review in the Particle Data Group's *Review of Particle Physics* [24], as well as the reviews by Marsh [468, 469], Kim [470], and Duffy and van Bibber [471].

Problems

17.1 Using the definition of R-parity in Eq. 17.3, show that $R = +1$ for all Standard Model particles and $R = -1$ for all of their superpartners.

17.2 One might think that the superpartner of the spin-1 gluon could be either a spin-1/2 or a spin-3/2 particle. In a viable supersymmetric theory, however, the number of degrees of freedom associated with the gluino has to be equal to that associated with the gluon, forcing us to pick the spin-1/2 option. Carry out this exercise for the entire MSSM, determining the spin of each of the superpartners.

17.3 Use the chargino mass matrix shown in Eq. 17.15 to confirm that the mass eigenvalues are those given in Eq. 17.16. Then perform an expansion in the $|M_2 \pm \mu| \gg m_W$ limit to verify the expressions given in Eq. 17.17.

17.4 Find the eigenvalues of the matrix given in Eq. 17.22. Then show that in the $\mu \ll M$ limit, these two masses are equal to μ^2/M and M.

17.5 The Boltzmann equation for the evolution of the number density of sterile neutrinos in the early universe is sometimes written as

$$\frac{dr}{d\ln a} = \frac{\gamma}{H} + r\,\frac{d\ln g_{\star,S}}{d\ln a}, \tag{17.75}$$

where $r \equiv n_s/n_a$ is the ratio of the number densities of sterile to active neutrinos, a is the scale factor, and $\gamma \equiv \sin^2 2\theta_{\rm eff}\,\Gamma/2$. Show that this is equivalent to the Boltzmann equation given in Eq. 17.29.

17.6 Show that if the effective mixing angle were constant, $\theta_{\rm eff} = \theta$, then the Boltzmann equation describing the evolution of the abundance of sterile neutrinos would be the same as in the case of UV-dominated freeze in (as shown in Eq. 9.47).

17.7 Show that the axion mass given in Eq. 17.43 follows from the potential in Eq. 17.42.

17.8 Consider an axion with $f_{\rm PQ}/N = 10^{10}$ GeV and $m_a = 5.7 \times 10^{-4}$ eV. Roughly estimate the contribution to $N_{\rm eff}$ from the thermal population of these particles.

17.9 Near the minima of the potential given in Eq. 17.42, show that $V(\phi) \approx (m_a^2/2)(f_{\rm PQ}/N)^2\,\theta_{\rm QCD}^2$.

17.10 Consider a scenario in which $f_{\rm PQ} = 10^{12}$ GeV. At the time of Peccei-Quinn symmetry breaking, how much energy was contained within the horizon? Compare this to the mass of a dwarf galaxy, and to the mass of the observable universe.

18

Direct Dark Matter Searches

Experiments designed to detect the interactions of dark matter particles with Standard Model targets allow us to test many well-motivated candidates for this mysterious substance. In particular, in many WIMP models, the dark matter is predicted to scatter with nuclei at potentially observable rates. In this chapter, I describe the kinematics of dark matter-nuclei scattering, and the calculation of the rates of these interactions. I also discuss how dark matter axions could be directly detected through their conversion into photons in the presence of a strong magnetic field.

18.1 Dark matter scattering with nuclei

Consider a dark matter particle that is moving through the halo with a momentum $\vec{p} = m_X \vec{v}$, and that then scatters elastically with a nucleus at rest. Applying momentum conservation, the kinetic energies of the dark matter particle before and after the scattering event are given by

$$E_i = \frac{\vec{p}^{\,2}}{2m_X}, \tag{18.1}$$

$$E_f = \frac{\vec{p}^{\,\prime 2}}{2m_X} = \frac{(\vec{p}-\vec{q})^2}{2m_X},$$

where $\vec{p}^{\,\prime}$ is the momentum of the dark matter particle after the collision, and $\vec{q}$ is the momentum of the nucleus after the collision (see Fig. 18.1). This latter momentum is related to the energy of the nuclear recoil by $E_R = \vec{q}^{\,2}/2m_N$. Applying energy conservation, we have

$$\frac{\vec{p}^{\,2}}{2m_X} = \frac{(\vec{p}-\vec{q})^2}{2m_X} + \frac{\vec{q}^{\,2}}{2m_N} \tag{18.2}$$

$$\longrightarrow \quad \frac{2(\vec{p}\cdot\vec{q}) - \vec{q}^{\,2}}{2m_X} = \frac{\vec{q}^{\,2}}{2m_N}$$

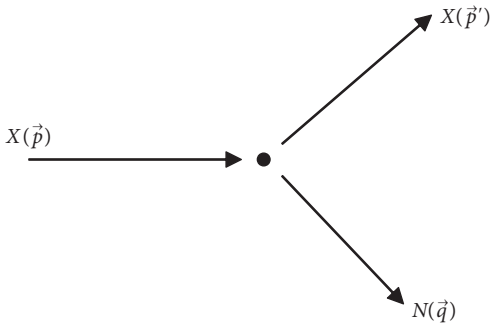


Figure 18.1 A dark matter particle, X, scattering with a nucleus, N.

$$\longrightarrow \quad \frac{\vec{p}\cdot\vec{q}}{m_X} = \frac{\vec{q}^{\,2}}{2}\left(\frac{1}{m_N}+\frac{1}{m_X}\right)$$

$$\longrightarrow \quad \frac{|\vec{p}\,||\vec{q}\,|\cos\theta}{m_X} = \frac{\vec{q}^{\,2}}{2\mu_{XN}},$$

where in the last line I have introduced the angle between $\vec{p}$ and $\vec{q}$, $\cos\theta = \vec{p}\cdot\vec{q}/(|\vec{p}\,||\vec{q}\,|)$, as well as the reduced mass of the dark matter-nucleus system, $\mu_{XN} = m_X m_N/(m_X + m_N)$. The recoil energy transferred to the nucleus is maximized for $\cos\theta = 1$, for which

$$\begin{aligned} E_R^{\max} &= \frac{|\vec{q}\,|^2_{\max}}{2m_N} \\ &= \frac{2\mu_{XN}^2\vec{p}^{\,2}}{m_N m_X^2} \\ &= \frac{2\mu_{XN}^2\vec{v}^{\,2}}{m_N}. \end{aligned} \tag{18.3}$$

For velocities of $v \sim 10^{-3}$—as are typical of dark matter particles in the halo of the Milky Way—and for relatively heavy dark matter particles, $m_X \gtrsim m_N$, this can correspond to recoil energies as large as $E_R^{\max} \sim 20 - 200$ keV for commonly used nuclear targets. For lighter dark matter particles, the maximum recoil energy imparted is further suppressed by an additional factor of $(m_X/m_N)^2$. Thus, in order for an experiment to detect the scattering of dark matter particles, it must be sensitive to very low-energy recoils.

We can calculate the scattering rate of dark matter particles with a target from the differential cross section for this process and the velocity distribution of the dark matter in the halo:

$$\frac{dR}{dE_R} = N_T\, n_X \int \frac{d\sigma_N}{dE_R}|\vec{v}|f(\vec{v})d^3\vec{v}, \tag{18.4}$$

where N_T is the number of nuclei in the target, n_X the local number density of dark matter particles, $d\sigma_N/dE_R$ is the dark matter's differential scattering cross section with a target nucleus, $\vec{v}$ is the velocity of a dark matter particle relative to the target, and $f(\vec{v})$ is the velocity distribution of the dark matter particles, normalized such that $\int f(\vec{v}\,)d^3\vec{v} = 1$.

For the velocities of the dark matter particles in the local halo of the Milky Way, $f(\vec{v})$ is often parameterized in terms of a Maxwellian distribution, truncated above the escape velocity of the galaxy. In the frame of the Milky Way, this is given by

$$f_{\rm MW}(\vec{v}) = C\, e^{-\vec{v}^{\,2}/v_0^2}\ \Theta(v_{\rm esc} - |\vec{v}|), \tag{18.5}$$

where $v_0 \approx 220\,{\rm km/s}$, $v_{\rm esc} \approx 550\,{\rm km/s}$, and the normalization constant is given by

$$C = \frac{1}{\pi^{3/2} v_0^3}\left[{\rm erf}\left(\frac{v_{\rm esc}}{v_0}\right) - \frac{2}{\sqrt{\pi}}\frac{v_{\rm esc}}{v_0} e^{-\vec{v}^{\,2}/v_0^2}\right]^{-1}. \tag{18.6}$$

Further taking into account the motion of the solar system through the Milky Way, we can write this velocity distribution as it appears in earth's rest frame:

$$f(\vec{v}) = C\, e^{-(\vec{v}+\vec{v}_\odot)^2/v_0^2}\ \Theta(v_{\rm esc} - |\vec{v} + \vec{v}_\odot|), \tag{18.7}$$

where $|\vec{v}_\odot| \approx 240\,{\rm km/s}$.

The cross section for a dark matter particle to scatter with a nucleus is a highly model-dependent quantity, and will be the subject of the following section.

18.2 Dark matter scattering cross sections

To evaluate the scattering cross section of a dark matter particle with a nucleus, we will first consider the interactions that take place with individual quarks and gluons. From these interactions, we will then infer how the dark matter interacts with nucleons, and how those interactions with nucleons lead to scattering with nuclei.

In many dark matter models, the amount of momentum that is transferred in scattering with a nucleus is much smaller than the mass of the particle that mediates these interactions. In such cases, it is possible to approach these interactions within the language of effective field theory, which greatly simplifies the dynamics at hand. While there are many different ways that the dark matter could potentially interact with nuclei, these interactions can in many cases be described in terms of effective scalar, vector, or axial couplings between the dark matter particle and quarks.

Let's first consider vector interactions, for which the squared amplitude takes the following form:

$$|\mathcal{M}|^2 = b_{p,n}^2\,(4 m_X m_{p,n})^2 \qquad \text{(vector coupling)}, \tag{18.8}$$

where the subscripts p and n denote protons and neutrons. In this case, the coefficients $b_{p,n}$ are directly related to the sum of the corresponding couplings between the dark matter and each of the three valance quarks in a given nucleon:

$$b_p = 2b_u + b_d = 2\frac{g_X g_u}{m_M^2} + \frac{g_X g_d}{m_M^2}, \tag{18.9}$$

$$b_n = b_u + 2b_d = \frac{g_X g_u}{m_M^2} + 2\frac{g_X g_d}{m_M^2},$$

where g_X and g_q are the couplings of the dark matter and the quarks to the particle that mediates the interaction, and m_M is the mass of that mediator.

Following Eq. 3.14, we can use the squared amplitude to obtain the dark matter's scattering cross section with nucleons:

$$\sigma_{p,n} = \frac{1}{4\sqrt{(E_i m_{p,n})^2 - (m_X m_{p,n})^2}} \int |\mathcal{M}|^2 (2\pi)^4 \delta(\vec{p} - \vec{p}\,' - \vec{q})\, \delta(E_i - E_f - E_R) \times \frac{1}{2m_X} \frac{d^3\vec{p}\,'}{(2\pi)^3} \frac{1}{2m_{p,n}} \frac{d^3\vec{q}}{(2\pi)^3}$$

$$= \frac{b_{p,n}^2}{4\pi^2 |\vec{v}|} \int \delta(\vec{p} - \vec{p}\,' - \vec{q})\, \delta(E_i - E_f - E_R)\, d^3\vec{p}\,'\, d^3\vec{q}$$

$$= \frac{b_{p,n}^2}{4\pi\, \vec{v}^{\,2}} \int \delta\left(\cos\theta - \frac{|\vec{q}|}{2\mu_{Xp,n} |\vec{v}|} \right) d\cos\theta\, d|\vec{q}|^2$$

$$= \frac{\mu_{Xp,n}^2\, b_{p,n}^2}{\pi} \qquad \text{(vector coupling).} \tag{18.10}$$

We can approach the scattering of a dark matter particle with a nucleus, N, in a similar fashion. In this case, however, we have to take into account the fact that the scattering cross section will be suppressed if the wavelength associated with the momentum transferred, $\vec{q}$, is not large compared to the size of the target nucleus. This is generally taken into account by including in the cross section the square of a nuclear form factor, F, which is a function of both the momentum transferred and the characteristics of the target nucleus. In the plane wave approximation, the form factor is the Fourier transform of the nucleon density within the target:

$$F(q) = \int d^3\vec{r}\, e^{i\vec{q}\cdot\vec{r}}\, n(r) \tag{18.11}$$

$$= \int_0^{2\pi} d\phi \int_0^{\infty} dr\, r^2 n(r) \int_{-1}^{+1} e^{iqr\cos\theta}\, d\cos\theta$$

$$= \frac{4\pi}{q} \int_0^{\infty} r \sin(qr)\, n(r) dr,$$

where $n(r)$ is the mass density of scattering centers within the target nucleus, and quantities without an arrow represent the magnitudes of the corresponding vectors (i.e., $q \equiv |\vec{q}|$).

The distribution of scattering centers depends on both the nucleus in question, and the nature of the interaction itself. In particular, so-called "spin-independent" interactions probe the distribution of nucleons throughout the nucleus, where as "spin-dependent" interactions couple to the net spin of a nucleus. In the spin-independent case, a reasonable approximation is to model the distribution of targets as a sphere of uniform density: $n(r) = A/(4\pi r_{\rm max}^3/3)$ for all $r < r_{\rm max}$, where A is the atomic mass of the nucleus and $r_{\rm max} \approx 1.2\, A^{1/3}$ fm. In this case, the form factor is given by

$$F_{\rm SI}(q) = \frac{4\pi}{q} \int_0^{r_{\rm max}} r \sin(qr)\, n\, dr, \tag{18.12}$$

$$= A\, \frac{3\,[\sin(qr_{\rm max}) - qr_{\rm max}\cos(qr_{\rm max})]}{(qr_{\rm max})^3} \qquad \text{(uniform sphere).}$$

Another parameterization that is commonly used in the spin-independent case is the so-called Helm form factor:

$$F_{\rm SI}(q) = A\,\frac{3\,[\sin(qr_{\rm max}) - qr_{\rm max}\cos(qr_{\rm max})]}{(qr_{\rm max})^3}\,e^{-(qs)^2/2} \qquad \text{(Helm)}, \tag{18.13}$$

where $s \approx 1$ fm.

Notice that if $q \ll r_{\rm max}^{-1}$, these form factors reduce to $F_{\rm SI} \approx A$, reflecting the total number of nucleons in the target. Thus, in interactions in which only a small amount of momentum is transferred, the dark matter particle can scatter coherently with the entire nucleus, resulting in a total amplitude that is equal to the sum of the amplitudes for the individual nucleons. This leads to a total scattering cross section that is proportional to the square of the atomic number, $\sigma \propto F^2 \propto A^2$. In contrast, if $q \gtrsim r_{\rm max}^{-1}$, corresponding to $E_R \gtrsim 7\,{\rm keV} \times (100/A)^{5/3}$, the de Broglie wavelength will be too small to cover the entire nucleus, limiting the degree to which the scattering cross section will benefit from this coherence.

With our form factor in hand, we can now generalize Eq. 18.10 to the case of a dark matter particle scattering with a nucleus. To accomplish this, we perform the following replacement:

$$\begin{aligned} b_{p,n}^2 &\longrightarrow \left[\sum_i b_{p,n}\right]^2 \left[\frac{F(q)}{A}\right]^2 \\ &= \Big[Z\,b_p + (A-Z)\,b_n\Big]^2 \left[\frac{F(q)}{A}\right]^2, \end{aligned} \tag{18.14}$$

where the sum is over the nucleons contained within the nucleus, and F is the form factor. Further replacing $d|\vec{q}|^2 = 2m_N dE_R$, we now have

$$\begin{aligned} \frac{d\sigma_N}{dE_R} &= \left[\frac{Z}{A}\,b_p + \frac{(A-Z)}{A}\,b_n\right]^2 \frac{m_N}{2\pi\,v^2}\,\delta\!\left(\cos\theta - \frac{|\vec{q}|}{2\mu_{XN}\,v}\right) F^2(q)\,d\cos\theta \\ &= \left[\frac{Z}{A}\,\frac{\sigma_p^{1/2}}{\mu_{Xp}} + \frac{(A-Z)}{A}\,\frac{\sigma_n^{1/2}}{\mu_{Xn}}\right]^2 \frac{m_N}{2v^2}\,\delta\!\left(\cos\theta - \frac{|\vec{q}|}{2\mu_{XN}\,v}\right) F^2(q)\,d\cos\theta. \end{aligned} \tag{18.15}$$

In integrating over $\cos\theta$, the delta function in this expression yields unity if $|\vec{q}| < 2\mu_{XN}\,v$, and zero otherwise. With this in mind, we can rewrite this in terms of a step function in velocity:

$$\frac{d\sigma_N}{dE_R} = \left[\frac{Z}{A}\,\frac{\sigma_p^{1/2}}{\mu_{Xp}} + \frac{(A-Z)}{A}\,\frac{\sigma_n^{1/2}}{\mu_{Xn}}\right]^2 \frac{m_N}{2v^2}\,F^2(q)\,\Theta\big(v - v_{\rm min}\big), \tag{18.16}$$

where $v_{\rm min} = (m_N E_R/2\mu_{XN}^2)^{1/2}$. We can now use this differential cross section in Eq. 18.4 to calculate the scattering rate at a given experiment per unit recoil energy.

Note that, for modest values of the recoil energy, the total cross section described by Eq. 18.16 is approximately proportional to the *square* of the number of nucleons in the target nucleus. This important feature has made it possible for direct detection experiments to place very stringent constraints on the spin-independent scattering of dark matter, in particular by using targets consisting of heavy elements, such as xenon or germanium.

Table 18.1 Matrix elements for the light quarks and gluons within a nucleon [472].

	Protons	Neutrons
$f_{T_u}^{p,n}$	0.018	0.013
$f_{T_d}^{p,n}$	0.027	0.040
$f_{T_s}^{p,n}$	0.037	0.037
$f_{TG}^{p,n}$	0.917	0.910
$\Delta_u^{p,n}$	0.80	−0.46
$\Delta_d^{p,n}$	−0.46	0.80
$\Delta_s^{p,n}$	−0.12	−0.12

For the next case, we will consider scattering that results from a scalar coupling between the dark matter and quarks. In contrast to the vector case, these couplings are generally proportional to the masses of the quarks. Thus, while we could safely focus on the light valence quarks in the vector case, we now have to consider the couplings of the dark matter to the heavier sea quarks, as well as those to gluons, which are induced through diagrams involving a loop of heavy quarks.

The nucleon-level cross section in this case can be written as follows:

$$\sigma_{p,n} = \frac{\mu_{Xp,n}^2 \lambda_X^2}{\pi m_M^4} f_{p,n}^2 \qquad \text{(scalar coupling)}, \tag{18.17}$$

where λ_X is the coupling of the dark matter to the mediator, m_M is the mediator's mass, and

$$\frac{f_{p,n}}{m_{p,n}} = \sum_{q=u,d,s} f_{T_q}^{p,n} \lambda_q + \frac{2}{27} f_{TG}^{p,n} \sum_{q=c,b,t} \lambda_q. \tag{18.18}$$

The quantity $m_q\,\lambda_q$ is the coupling of the mediator to a quark of mass m_q, and the quantities $f_{T_q}^{p,n}$ and $f_{TG}^{p,n}$ are matrix elements that can be found in Table 18.1. The first term in this expression accounts for the scattering of the dark matter with quarks in the nucleon, while the second term corresponds to the scattering with gluons through a heavy quark loop. In the case of a mediator that couples universally to all flavors of quarks ($\lambda_q = \lambda$ for all q), the total effective coupling reduces to $f_{p,n} = \lambda\, m_{p,n}[f_{T_u}^{p,n} + f_{T_d}^{p,n} + f_{T_s}^{p,n} + (6/27) f_{TG}^{p,n}] \approx 0.29 \times \lambda\, m_{p,n}$. By inserting the nucleon-level cross section given in Eq. 18.17 into Eq. 18.16, we can obtain the dark matter's scattering cross section with nuclei for the case of a scalar coupling.

As a third case, we can consider axial interactions between the dark matter and quarks. The nucleon-level cross section in this case can be written as follows:

$$\sigma_{p,n} = \frac{3\mu_{Xp,n}^2 g_X^2}{\pi m_M^4} \left[\sum_{q=u,d,s} \Delta_q^{p,n} g_q \right]^2 \qquad \text{(axial coupling)}, \tag{18.19}$$

where g_X and g_q are the couplings of the mediator to the dark matter and quarks, and m_M is the mediator's mass. The quantities $\Delta_q^{p,n}$ are related to the spin that is carried by the various quarks within a nucleon, with values as given in Table 18.1.

Since axial interactions couple the dark matter to the spin content of the target (rather than to its total mass), the form factor is different in this case. Taking the correct spin-dependent form factor into account, the differential scattering cross section that results from this coupling (analogous to Eq. 18.16) is given by:

$$\frac{d\sigma_N}{dE_R} = \left[\frac{\langle S_p\rangle}{J}\frac{\sigma_p^{1/2}}{\mu_{Xp}} + \frac{\langle S_n\rangle}{J}\frac{\sigma_n^{1/2}}{\mu_{Xn}}\right]^2 \frac{m_N}{2v^2}\, F^2(q)\, \Theta\big(v - v_{\rm min}\big), \tag{18.20}$$

where $\langle S_{p,n}\rangle$ are the expectation values of the spin carried by the protons and neutrons in the target nucleus, and J is the total spin of that nucleus. At low to moderate recoil energies, $F^2(q) \approx J(J+1)$. In contrast to spin-independent scattering, this cross section is not necessarily larger for heavier nuclei.

18.3 Direct detection constraints and their impact

The idea of detecting dark matter particles through their scattering with nuclei goes back to the mid-1980s [473, 474]. Only a few years later, the first experimental limits on the dark matter's cross section with nuclei had been obtained [475, 476]. In recent decades, the constraints from this class of experiments have improved at an exponential rate, increasing in sensitivity by a factor of 2 every 14 months or so. The most recent of these results are summarized in Figure 18.2. In the case of spin-independent scattering, the strongest constraints have been placed by the LUX-ZEPLIN (LZ) experiment [477], as well as by CRESST-III [478], CDMSlite [479], DarkSide-50 [480], and XENON1T [481, 482]. The results shown in the top frame of Figure 18.2 assume a common cross section for protons and neutrons, $\sigma_p = \sigma_n$. In the case of spin-dependent scattering, the strongest constraints have been placed by PICO [483] (for scattering with protons) and LZ [484] (for neutrons).

In each frame of the figure, I also show the cross section at which solar and atmospheric neutrinos are predicted to become a significant background for this class of experiments [485, 486]. This background is expected to make it difficult for direct detection experiments to test dark matter models that feature cross sections below this so-called "neutrino floor." In the case of spin-independent scattering, the location of the neutrino floor is largely insensitive to the type of target nucleus employed. In contrast, the neutrino floor for spin-dependent scattering depends more strongly on the nuclear species under consideration. In Figure 18.2, I show the neutrino floor for the case of a xenon target and for dark matter scattering with neutrons [486].

To better understand the impact of these constraints, let's look at a few examples. As a first case, consider a Dirac fermion that annihilates through the exchange of a heavy mediator ($m_M \gg 2m_X$) with vector couplings to the dark matter particle and to all Standard Model quarks. The nonrelativistic annihilation cross section for this dark matter candidate is given by

$$\sigma\, v \approx \frac{3 n_q g_X^2 g_q^2 m_X^2}{\pi\, m_M^4} \qquad \text{(vector coupling)}, \tag{18.21}$$

where I have included a factor of n_q for the number of quark species that the dark matter can annihilate to without suppression (i.e., those with $m_X \gg m_q$). In order for this particle to freeze out of thermal equilibrium with an abundance equal to the measured dark matter density, we need $\sigma\, v \approx 2.2 \times 10^{-26}\ {\rm cm}^3/{\rm s}$ (see Sec. 9.5.2), thus requiring $g_X g_q / m_M^2 \approx$

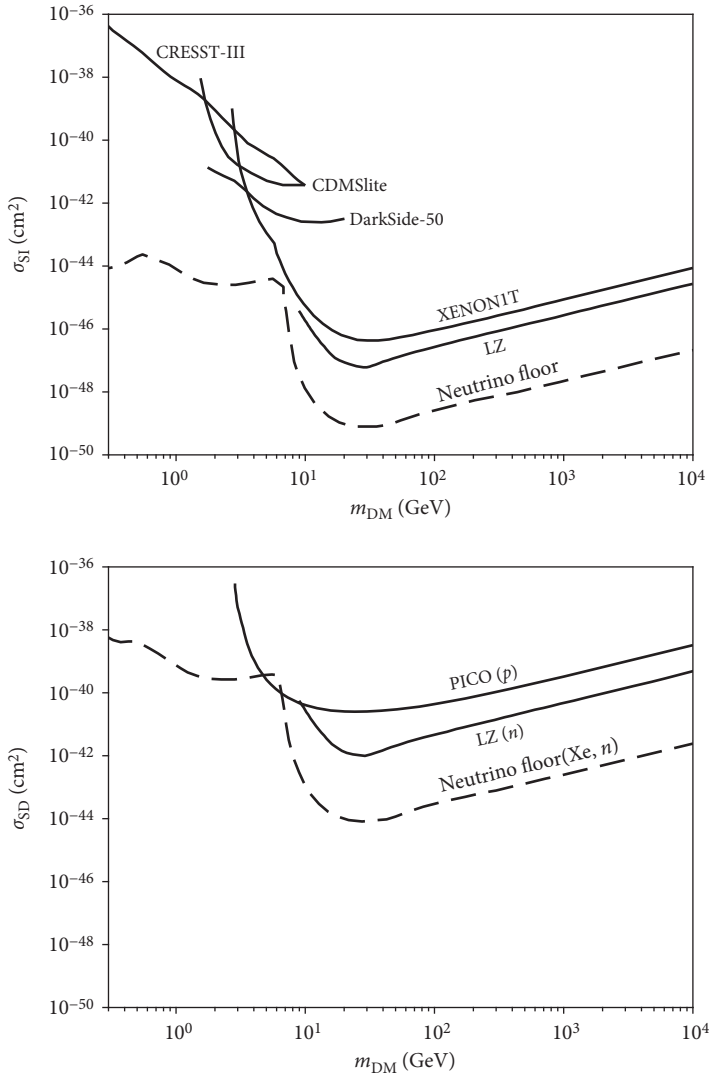


Figure 18.2 Constraints on the spin-independent (top) and spin-dependent (bottom) scattering cross sections of dark matter with nucleons. Also shown is the so-called "neutrino floor," at which dark matter experiments are expected to encounter significant backgrounds from neutrinos.

$4.4 \times 10^{-5}\,\mathrm{GeV}^{-1}/(m_X\, n_q^{1/2})$. These couplings also lead to the following spin-independent scattering cross section with nucleons:

$$\begin{aligned}\sigma_{p,n} &\approx \frac{9\mu_{Xp,n}^2 g_X^2 g_q^2}{\pi m_M^4} \qquad (18.22)\\ &\approx \frac{9\mu_{Xp,n}^2}{\pi}\left(\frac{4.4\times 10^{-5}\,\mathrm{GeV}^{-1}}{m_X n_q^{1/2}}\right)^2\\ &\approx 4\times 10^{-41}\,\mathrm{cm}^2 \times \left(\frac{100\,\mathrm{GeV}}{m_X}\right)^2\left(\frac{5}{n_q}\right) \qquad \text{(vector coupling)},\end{aligned}$$

where in the second line I used the value of $g_X g_q / m_M^2$ that we found was needed to obtain the measured dark matter abundance. In the last line, I also assumed that $m_X \gg m_{p,n}$. Comparing this to the constraint on spin-independent scattering shown in the top frame of Figure 18.2, we see that this dark matter candidate is ruled out for a wide range of masses, from $m_X \sim 0.3\,\text{GeV}$ up to $\sim 12\,\text{TeV}$.

Repeating this exercise for dark matter interactions that are mediated by a heavy scalar, the annihilation cross section in this case is given by

$$\sigma v \approx \frac{3 v^2 \lambda_X^2 m_X^2}{8\pi m_M^4} \sum_q \lambda_q^2 m_q^2 \qquad \text{(scalar coupling)}. \tag{18.23}$$

In this scenario, we expect the annihilation cross section to be dominated by the couplings to the heaviest quark species that the dark matter can annihilate into, allowing us to replace $\sum_q \lambda_q^2 m_q^2$ with $\lambda_q^2 m_q^2$ for the heaviest quark with $m_q \ll m_X$. Using $v^2 \approx 0.1$ (the approximate value at the time of dark matter freeze out), we find that $\lambda_X \lambda_q / m_M^2 \approx 3.9 \times 10^{-4}\,\text{GeV}^{-1} / m_X m_q$ is required to obtain the measured dark matter density. Once again, we can use this combination of quantities to calculate the dark matter's spin-independent scattering cross section with nucleons:

$$\begin{aligned}
\sigma_{p,n} &\approx \frac{\mu_{Xp,n}^2 \lambda_X^2 m_{p,n}^2}{\pi m_M^4} \left[\sum_{q=u,d,s} f_{T_q}^{p,n} \left(\frac{\lambda_q}{\text{GeV}} \right) + \frac{2}{27} f_{TG}^{p,n} \sum_{q=c,b,t} \left(\frac{\lambda_q}{\text{GeV}} \right) \right]^2 \qquad (18.24) \\
&\approx \frac{\mu_{Xp,n}^2 m_{p,n}^2}{\pi} \left(\frac{3.9 \times 10^{-4}\,\text{GeV}^{-1}}{m_X m_q} \right)^2 \left[\sum_{q=u,d,s} f_{T_q}^{p,n} + \frac{6}{27} f_{TG}^{p,n} \right]^2 \\
&\approx 7 \times 10^{-42}\,\text{cm}^2 \times \left(\frac{100\,\text{GeV}}{m_X} \right)^2 \left(\frac{4.2\,\text{GeV}}{m_q} \right)^2 \qquad \text{(scalar coupling)},
\end{aligned}$$

where I have scaled this result to the mass of the bottom quark, $m_b \approx 4.2\,\text{GeV}$. This case is ruled out for $\text{GeV} \lesssim m_X \lesssim 6\,\text{TeV}$. If the dark matter candidate is heavy enough to annihilate into a pair of top quarks, however, the relatively small value of λ_t that is required to obtain the measured dark matter abundance leads to a more modest elastic scattering rate,

$$\sigma_{p,n} \approx 1.6 \times 10^{-45}\,\text{cm}^2 \times \left(\frac{500\,\text{GeV}}{m_X} \right)^2 \left(\frac{173\,\text{GeV}}{m_q} \right)^2 \qquad \text{(scalar coupling)}, \tag{18.25}$$

which is below current constraints for $m_X \gtrsim 500\,\text{GeV}$.

Lastly, we will consider dark matter that interacts with quarks through a mediator that has purely axial couplings. The annihilation cross section in this case is given by:

$$\sigma v \approx \frac{3 g_X^2}{2\pi m_M^4} \sum_q g_q^2 \left[m_q^2 - \frac{m_X^2}{3} v^2 \right] \qquad \text{(axial coupling)}. \tag{18.26}$$

If the first term in the square brackets (the *s*-wave term) dominates, the dark matter will annihilate mostly into the heaviest kinematically accessible quark species. To obtain the measured dark matter density in this case, we require $g_X g_q / m_M^2 \approx 6.2 \times 10^{-5}\,\text{GeV}^{-1} / m_q$,

leading to the following spin-dependent scattering cross section with nucleons:

$$\sigma_{p,n} = \frac{3\mu_{Xp,n}^2 g_X^2}{\pi m_M^4}\left[\sum_{q=u,d,s} \Delta_q^{p,n} g_q\right]^2 \tag{18.27}$$

$$\approx \frac{3\mu_{Xp,n}^2}{\pi}\left(\frac{6.2\times10^{-5}\,\mathrm{GeV}^{-1}}{m_q}\right)^2\left[\sum_{q=u,d,s} \Delta_q^{p,n}\right]^2$$

$$\approx 2\times10^{-42}\,\mathrm{cm}^2 \times \left(\frac{173\,\mathrm{GeV}}{m_q}\right)^2 \qquad \text{(axial coupling)}.$$

Comparing this to the lower frame of Figure 18.2, we conclude that this scenario is consistent with current constraints so long as the dark matter is heavy enough to efficiently annihilate to top quarks.

While the constraints from direct detection experiments do rule out many otherwise viable dark matter candidates, they also leave many other WIMP models untested. Furthermore, there are many ways in which the dark matter's elastic scattering cross section with nuclei could be suppressed relative to the estimates presented in this section. First of all, if the dark matter annihilates mostly to gauge or Higgs bosons, or to leptons, it would require smaller couplings to quarks and thus scatter with nuclei at a lower rate. This is the case for both wino-like and higgsino-like neutralinos, for example [381, 382, 375] (see Sec. 17.1.3). Second, in many models, the dark matter's elastic scattering cross section with nuclei is further suppressed by multiple powers of velocity or momentum, significantly reducing their scattering rates. Lastly, if the dark matter annihilates near a resonance, $2m_X \sim m_M$, smaller couplings will be required, again leading to suppressed cross sections with nuclei.

18.4 Axion detection

Axions are far too light and feebly interacting to be detected through their elastic scattering with nuclei. Instead, efforts to directly detect axions in the local dark matter halo generally involve experiments that are designed to convert these particles into photons in the presence of a strong magnetic field [415, 487, 488]. The idea here is to tune the geometry of the cavity containing the magnetic field in such a way that incoming dark matter axions will be resonantly converted into radio-frequency photons, $h\nu = m_a c^2\,(1+\mathcal{O}(v^2\sim10^{-6}))$. The power in photons that is generated through axion conversion in such an experiment is given by

$$P_a = g_{a\gamma\gamma}^2\,\frac{VB^2\rho_a Q}{m_a}, \tag{18.28}$$

$$= \left(\frac{\alpha g_\gamma}{\pi\,(f_{\rm PQ}/N)}\right)^2 \frac{VB^2\rho_a Q}{m_a}$$

$$\approx 6\times10^{-19}\,\mathrm{W}\times\left(\frac{g_\gamma}{-0.97}\right)^2\left(\frac{2.85\times10^{11}}{f_{\rm PQ}/N}\right)^2\left(\frac{m_a}{2\times10^{-5}\,\mathrm{eV}}\right)\left(\frac{V}{0.1\,\mathrm{m}^3}\right)$$

$$\times\left(\frac{B}{10^5\,\mathrm{G}}\right)^2\left(\frac{\rho_a}{0.4\,\mathrm{GeV/cm}^3}\right)\left(\frac{Q}{10^5}\right),$$

where $g_{a\gamma\gamma}$ is the dimensionful coupling of the axion to two photons (see Eq. 17.45), V is the volume of the cavity, B is the strength of the magnetic field, ρ_a is the local density of axions, and Q is the quality factor of the cavity.[1]

This amount of power is tiny for QCD axions in the mass range favored by cosmological considerations. Furthermore, the frequency of a cavity must be precisely tuned to search for axions of a given mass. To hope to detect axions with such an experiment, one must operate at a given frequency long enough to achieve a detectable signal-to-noise ratio:

$$\begin{aligned}\frac{\text{Signal}}{\text{Noise}} &= \frac{P_a\, t}{T_n \sqrt{t\, \Delta\nu}} \\ &\approx 1.9 \times \left(\frac{g_\gamma}{-0.97}\right)^2 \left(\frac{2.85 \times 10^{11}}{f_{\rm PQ}/N}\right)^2 \left(\frac{m_a}{2 \times 10^{-5}\,\text{eV}}\right)^{1/2} \left(\frac{V}{0.1\,\text{m}^3}\right) \left(\frac{B}{10^5\,\text{G}}\right)^2 \\ &\quad \times \left(\frac{\rho_a}{0.4\,\text{GeV/cm}^3}\right) \left(\frac{Q}{10^5}\right) \left(\frac{t}{100\,\text{sec}}\right)^{1/2} \left(\frac{10^{-6}}{\Delta\nu/m_a}\right)^{1/2} \left(\frac{0.5\,\text{K}}{T_n}\right),\end{aligned} \tag{18.29}$$

where t is the length of time that the experiment operates at a given frequency, $\Delta\nu \approx v^2 m_a/2\pi \sim 10^{-6}\, m_a/2\pi$ is the width of the axion signal, and T_n is the effective temperature of the noise associated with the experiment.

In recent years, the Axion Dark Matter Experiment (ADMX) has operated at a series of frequencies between 0.6 and 1 GHz, probing QCD axions with masses in the range of $m_a \approx (2.6 - 4.2) \times 10^{-6}$ eV [489, 490, 491]. Plans are underway to substantially expand the reach of this experimental program, with the goal of ultimately testing QCD axions over the entire mass range favored by cosmological considerations.

Additional reading

For a review of direct detection, I recommend Lin's TASI lectures [74], as well as those by Schumann [492] and Baudis [493].

Problems

18.1 Show that the spin-independent form factors given in Eqs. 18.12 and 18.13 reduce to A in the $q \ll r_{\rm max}^{-1}$ limit.

18.2 How would the spin-independent form factor be altered if nuclei were 10 times larger in radius? How would this impact our ability to detect dark matter's spin-independent scattering with xenon nuclei? Note that typical xenon detectors can only detect events with recoil energies greater than a few keV.

18.3 Consider a dark matter candidate that interacts through a spin-1 mediator with vector couplings to third generation fermions, but that does not couple to first or second generation fermions. Estimate this dark matter candidate's elastic scattering cross section with nuclei. For what range of masses is this dark matter candidate consistent with current constraints? Assume that $m_X \gg m_b$.

[1] A useful unit conversion for the strength of a magnetic field is $1\,G = 1\,\text{g}^{1/2}\,\text{cm}^{-1/2}\,\text{s}^{-1}$.

18.4 If the dark matter interacts through a spin-1 mediator that has vector couplings to the dark matter and axial couplings to quarks, the dark matter's elastic scattering cross section with nuclei will be approximately that given in Eq. 18.10, but with an additional factor of v^2. For velocities typical of dark matter particles in the local halo, $v \sim 10^{-3}$, estimate to what degree this dark matter candidate is currently constrained by direct detection experiments.

18.5 For dark matter with vector, scalar, or axial couplings to quarks, estimate the range of masses that could be tested by a direct detection experiment with a sensitivity near the level of the neutrino floor.

18.6 Dark matter in the form of a higgsino-like neutralino cannot scatter (at tree level) with nuclei. This particle can, however, scatter inelastically, being converted in the process into a slightly heavier state [494, 375]. Applying energy and momentum conservation, what is the minimum velocity that the incoming dark matter particle must have in order to upscatter off a xenon nucleus into a state that is 100 keV more massive? Take the mass of the higgsino to be 1 TeV.

18.7 For an axion cavity experiment with $V = 0.1\ \text{m}^3$, $B = 10^5$ G, $Q = 10^5$, and $T_n = 0.5$ K, estimate how long it would take to perform a scan that can achieve $S/N = 2$ for DFSZ axions across a range of masses between 10^{-6} and 10^{-5} eV in steps of $\Delta\nu/m_a = 10^{-5}$.

19

Searches for Dark Matter Annihilation and Decay

Many experiments and telescopes have been deployed in an effort to detect the products of dark matter annihilation or decay. These so-called "indirect searches" make use of both ground- and space-based detectors, capable of detecting gamma rays, antiprotons, positrons, and other particles that might be generated through these processes. This chapter describes some of the most important indirect searches for dark matter, including efforts to detect gamma rays from the Galactic Center and from dwarf galaxies, as well as endeavors to identify signals of dark matter annihilation or decay in the spectrum of cosmic-ray positrons, antiprotons, and antinuclei.

19.1 Motivations for indirect searches

Direct detection experiments have produced very stringent constraints, allowing us to rule out many otherwise attractive candidates for the dark matter of our universe. The elastic scattering cross section of a dark matter particle with a nucleus, however, is a highly model-dependent quantity. Even among dark matter candidates classified as WIMPs, the rates predicted at such experiments can vary by many orders of magnitude, depending on the detailed interactions of the particle in question.

In this respect, indirect searches have an advantage. For dark matter candidates that were produced through thermal freeze out in the early universe, their annihilation cross section is closely connected to their relic abundance. More specifically, recall from Sec. 9.5 that in order for a particle species to freeze out of equilibrium with an abundance equal to the measured density of dark matter, it must have an annihilation cross section of approximately $\langle \sigma v \rangle \approx 2.2 \times 10^{-26}$ cm^3/s, as evaluated at the temperature of freeze out. In many dark matter models, the annihilation cross section in the modern universe is similar to its value at the time of thermal freeze out, providing us with an important benchmark for indirect searches.

There are, of course, many possible caveats to the logic put forth in the previous paragraph. For example, the dark matter's annihilation cross section could be velocity-dependent, and thus could be larger or smaller in the current, low-temperature universe than it was at the time of thermal freeze out. Similarly, if factors such as coannihilations or a matter-antimatter asymmetry played an important role in the process of freeze out, the

dark matter's annihilation rate in the universe today could be significantly smaller than the benchmark value of $\langle \sigma v \rangle \approx 2.2 \times 10^{-26}$ cm^3/s.

19.2 Dark matter annihilation and the CMB

Before turning our attention to dark matter particles annihilating in the halo of the Milky Way, let's consider the impact that dark matter annihilation may have had on our universe during the era of recombination. The fraction of dark matter particles that annihilate in a given Hubble time can be estimated by the ratio of the annihilation rate to the expansion rate, $n_X \langle \sigma v \rangle / H$. Using the facts that $n_X \propto g_{\star,S}\, T^3$ and $H \propto g_\star^{1/2}\, T^2$ during radiation domination, and that $H \propto T^{3/2}$ during matter domination, we can evaluate this ratio as a function of temperature. In the matter-dominated era, and assuming that $\langle \sigma v \rangle$ is constant, this is given by

$$\begin{aligned} \frac{n_X \langle \sigma v \rangle}{H} &\approx \left(\frac{g_{\star,S}(T)\, T^3}{g_{\star,S}(T_f)\, T_f^3} \right) \left(\frac{g_\star^{1/2}(T_f)\, T_f^2}{g_\star^{1/2}(T_{\rm Eq})\, T_{\rm Eq}^2} \right) \left(\frac{T_{\rm Eq}^{3/2}}{T^{3/2}} \right) \\ &= \frac{T^{3/2}}{T_f\, T_{\rm Eq}^{1/2}} \frac{g_{\star,S}(T)}{g_\star^{1/2}(T_f)\, g_\star^{1/2}(T_{\rm Eq})}, \end{aligned} \tag{19.1}$$

where the subscripts denote the temperatures of freeze out and matter-radiation equality. Note that in the second line I have taken $g_\star(T_f) = g_{\star,S}(T_f)$.

For concreteness, consider a dark matter candidate with a mass of 100 GeV that froze out when the temperature was $T_f \approx 100\,{\rm GeV}/26 \approx 4$ GeV. From the expression given above, we estimate that $n_X \langle \sigma v \rangle / H \approx 8.7 \times 10^{-12} \times (T/0.26\,{\rm eV})^{3/2}$. This corresponds to the following rate of energy injection per Hubble time, per baryon:

$$\frac{n_X \langle \sigma v \rangle}{H} \times \frac{\Omega_{\rm DM}}{\Omega_b} m_p \approx 0.04\,{\rm eV} \times \left(\frac{T}{0.26\,{\rm eV}} \right)^{3/2}. \tag{19.2}$$

In other words, the process of dark matter annihilation injected on the order of ~ 0.04 eV into the universe for each baryon during the era of recombination (for this choice of parameters). This is enough energy to ionize up to $\sim (0.04\,{\rm eV})/(13.7\,{\rm eV}) \sim 10^{-3}$ of the hydrogen atoms, substantially impacting the process of recombination, as well as the observed characteristics of the CMB. In fact, measurements of the CMB by the Planck Collaboration have allowed us to place quite stringent constraints on the dark matter's annihilation cross section [8], at a level of

$$\frac{\langle \sigma v \rangle f_{\rm eff}}{m_X} < 3.2 \times 10^{-28}\,{\rm cm^3\,s^{-1}\,GeV^{-1}}, \tag{19.3}$$

where $\langle \sigma v \rangle$ is the value of the annihilation cross section during the era of recombination, and $f_{\rm eff}$ is the fraction of the energy injected through annihilations that is transferred into the intergalactic medium [495, 496, 497, 498, 499]. For purely electromagnetic annihilation channels, such as $XX \to e^+ e^-$, the efficiency is quite high, $f_{\rm eff} \sim 0.4 - 0.6$, leading to a constraint of $m_X \gtrsim 30\,{\rm GeV}$ for a velocity-independent thermal relic. For dark matter particles that predominantly annihilate into quarks, gauge bosons, Higgs bosons, or heavy

leptons, $f_{\rm eff} \sim 0.2$, leading to a somewhat weaker constraint, $m_X \gtrsim 10 - 15\,{\rm GeV}$. If the dark matter is a thermal relic, it can be lighter than this only if its annihilation cross section is temperature-dependent, or if its annihilation products do not efficiently heat or ionize the intergalactic medium, such as in the case of $XX \to \nu\bar{\nu}$.

19.3 Gamma rays from dark matter annihilation

If the dark matter's annihilation cross section is near the thermal relic benchmark value, $\langle \sigma v \rangle \simeq 2.2 \times 10^{-26}$ cm^3/s, their annihilations could potentially lead to observable fluxes of energetic particles, including gamma rays [500, 501]. Searches for dark matter using gamma-ray telescopes benefit from the fact that these particles are not deflected by magnetic fields and are negligibly attenuated over galactic distance scales, making it possible to acquire both spectral and spatial information, uncorrupted by astrophysical effects.

Searches for gamma rays from annihilating dark matter are conducted using a combination of space-based and ground-based telescopes, each of which offer various advantages and disadvantages. At energies between approximately 0.1 and 100 GeV, this field is dominated by the Fermi Gamma-Ray Space Telescope, which observes the entire sky with an angular resolution on the order of a degree and an energy resolution of around 10%. Fermi was designed to accomplish many goals, among them to offer unprecedented sensitivity to dark matter annihilating in the Galactic Center and in Milky Way dwarf galaxies. At higher energies, ground-based air Cherenkov telescopes offer the greatest sensitivity, including the Cherenkov Telescope Array (CTA) [502], as well as HESS [503], VERITAS [504], and MAGIC [505]. While these instruments have greater angular resolution than Fermi, they must be pointed at specific targets and are only sensitive to gamma rays with energies above several tens of GeV. Ground-based water Cherenkov telescopes, such as HAWC and LHAASO, are also sensitive to annihilating dark matter, in particular in the case of very heavy particles [506, 507].

The spectrum of gamma rays that is generated by annihilating dark matter depends on the mass of the dark matter particle and on the kinds of particles that the annihilations produce. Annihilations to quarks, for example, typically lead to the formation of QCD jets, which fragment and hadronize to create a number of pions, which decay to produce gamma rays. This generally results in a gamma ray spectrum that peaks at around $E_\gamma \sim m_X/20$ (in $E_\gamma^2\, dN_\gamma/dE_\gamma$ units). A similar spectrum results from dark matter particles that annihilate to combinations of gauge or Higgs bosons. In contrast, dark matter candidates that annihilate mostly to charged leptons can generate gamma rays with a rather different spectral shape. Annihilations to $\tau^+\tau^-$, for example, produce a harder spectrum of pions, leading to a gamma-ray spectrum that is fairly sharply peaked at around $\sim m_X/3$. In the case of annihilations to e^+e^- or $\mu^+\mu^-$, the gamma-ray spectrum is dominated by final state radiation (rather than pion decay), generally resulting in a smaller flux of high-energy photons. In practice, the detailed gamma-ray spectrum from annihilating dark matter is often calculated using event generating software, such as PYTHIA [508]. In addition to those gamma rays that the dark matter annihilations produce directly, any electrons and positrons that are generated through this process can go on to create secondary gamma rays through inverse Compton scattering and bremsstrahlung [509, 510, 511] (see Chapter 5).

The dark matter annihilation rate per volume is given by

$$\frac{d\Gamma_{XX}}{dV} = \frac{\langle \sigma v \rangle\, \rho_X^2}{2m_X^2}, \tag{19.4}$$

where ρ_X is the dark matter density and the factor of 1/2 is included to avoid double counting the annihilation of particle A with particle B and the annihilation of particle B with particle A. In this expression, it is assumed that the annihilating particles are their own antiparticle (XX). If we were to instead consider annihilations between a dark matter particle and its antiparticle ($X\bar{X}$), the annihilation rate would be half as large for a given value of the cross section. However, the annihilation cross section must also be twice as large in this case in order to obtain the desired relic abundance, and thus the overall annihilation rate of a thermal relic today remains the same, regardless of whether the dark matter candidate is or is not its own antiparticle.

To calculate the flux of gamma rays from a given region of the sky, we must integrate the annihilation rate and the spectrum produced per annihilation over the volume of the region observed:

$$\begin{aligned}\frac{dN_\gamma}{dE_\gamma} &= \int \frac{\langle \sigma\, v\rangle\, \rho_X^2}{2m_X^2} \left.\frac{dN_\gamma}{dE_\gamma}\right|_{\rm ann} \frac{dV}{4\pi\, d^2} \\ &= \frac{\langle \sigma\, v\rangle}{8\pi\, m_X^2} \left.\frac{dN_\gamma}{dE_\gamma}\right|_{\rm ann} \int_{\Delta\Omega}\int_{los} \rho_X^2(l,\Omega)\, dl\, d\Omega,\end{aligned} \tag{19.5}$$

where dN_γ/dE_γ with the subscript "ann" represents the spectrum of gamma rays produced per annihilation, and d is the distance from the observer. In the second line, the integrals are performed over a narrow cone of solid angle, $\Delta\Omega$, and over the line of sight, l. Note that the d^2 in the denominator of the first line cancels with the l^2 factor in $dV = l^2\, dl\, d\Omega$.

The quantity described by the double integral in the second line of Eq. 19.5 is known as the annihilation J-factor, which encodes all of the relevant information regarding the distribution of the dark matter. To build some intuition for the annihilation J-factor, consider the simple example of dark matter annihilating in a spherical volume of radius r with a uniform density ρ that is located at a distance d. This J-factor is given by

$$J(\Delta\Omega) \equiv \int_{\Delta\Omega}\int_{los} \rho_X^2(l,\Omega)\, dl\, d\Omega = \frac{4\pi\, r^3 \rho^2}{3d^2}. \tag{19.6}$$

From this simple example, we can see that the overall magnitude of the signal from annihilating dark matter will be the largest in the direction of those regions that have a high density of dark matter ($J \propto \rho^2$), that are nearby ($J \propto d^{-2}$), and that are extended across a large volume of space ($J \propto V$).

19.3.1 The Galactic Center

Numerical simulations of cold, collisionless dark matter find that these particles form halos with large central densities [512, 513]. A common parameterization for the dark matter distribution within a given halo is the generalized Navarro-Frenk-White (NFW) profile [514, 515]:

$$\rho(r) \propto \frac{(r/R_s)^{-\gamma}}{(1 + r/R_s)^{3-\gamma}}, \tag{19.7}$$

where r is the distance from the center of the halo and R_s is the halo's scale radius. In the case of the Milky Way's halo, $R_s \sim 20$ kpc. While the canonical NFW profile is defined such that $\gamma = 1$, other values for the inner slope are also commonly adopted (as well as other parameterizations, such as the Einasto profile [516]). In particular, modern simulations that

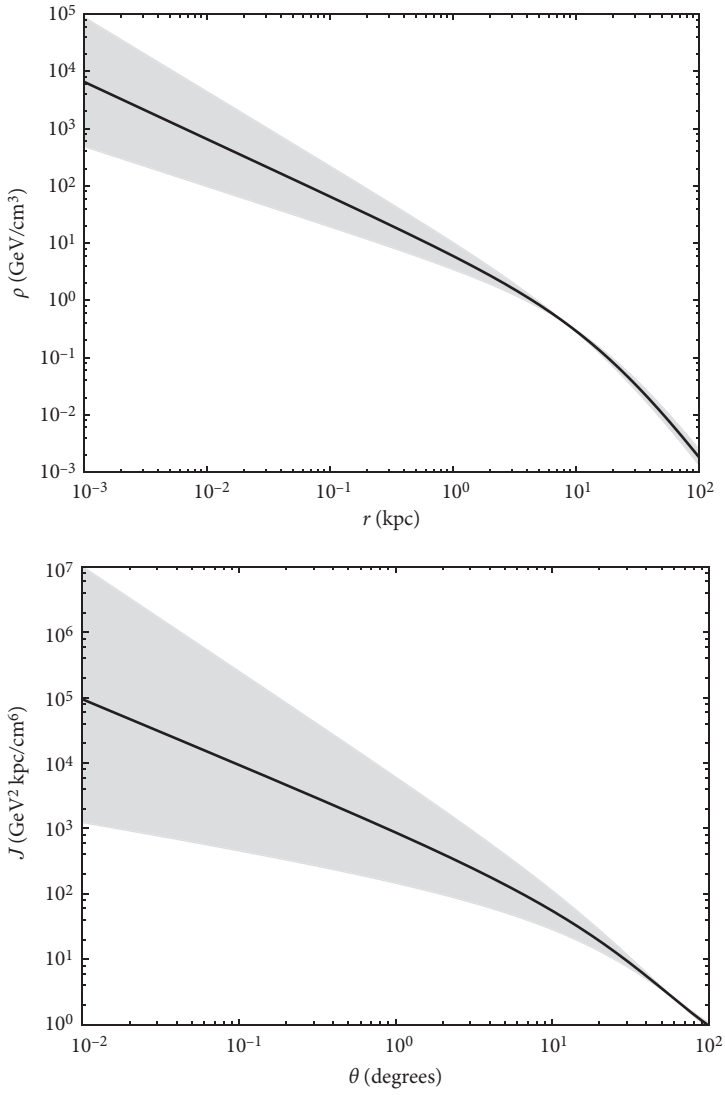


Figure 19.1 The generalized NFW profile for the Milky Way as a function of the distance to the Galactic Center (top), and the corresponding annihilation J-factors, as a function of the angle from the Galactic Center (bottom). The black curves are for the case of $\gamma = 1$, while the grey band corresponds to $\gamma = 0.7$ to 1.3.

include the effects of baryonic processes have been found to yield dark matter halos with a fairly wide range of inner profiles, $\gamma \sim 0.7 - 1.4$ [517, 518, 519, 520, 521, 522, 523, 524, 525, 526], and, empirically speaking, we have only limited information regarding the detailed distribution of dark matter in the Milky Way, especially in the kiloparsecs immediately surrounding the Galactic Center.

Regardless of the precise value of γ, the largest overall J-factors are expected to be found in the direction of the Galactic Center, making it the single brightest source of dark matter annihilation products on the sky. This direction, however, is also plagued by large

and imperfectly understood astrophysical backgrounds. At the other extreme are the Milky Way's dwarf galaxies, which have much smaller J-factors than the Galactic Center, but also feature much smaller gamma-ray backgrounds.

As an example, consider dark matter that is distributed according to an NFW profile with $\gamma = 1$ and $R_s = 20$ kpc, and normalized such that the local density of dark matter (at $r = 8.25$ kpc) is 0.4 GeV/cm^3. Numerically integrating Eq. 19.5 over the line of sight and over an angular region that encompasses the innermost 10° around the Galactic Center, we obtain the following flux of gamma-ray annihilation products:

$$F_\gamma \approx 3.6 \times 10^{-8}\,\mathrm{cm^{-2} s^{-1}} \times \left(\frac{\langle \sigma\, v\rangle}{2.2 \times 10^{-26}\,\mathrm{cm^3/s}}\right)\left(\frac{\int \frac{dN_\gamma}{dE_\gamma}\big|_{\rm ann} dE_\gamma}{10}\right)\left(\frac{100\,\mathrm{GeV}}{m_X}\right)^2. \tag{19.8}$$

For the case of the Fermi Gamma-Ray Space Telescope (with an effective area of $A \sim 8{,}500\,\mathrm{cm}^2$ and a field of view covering $\sim 20\%$ of this sky at a given time), this flux corresponds to a rate of approximately $\sim 2{,}000$ events per year. For a halo profile with an inner slope of $\gamma = 1.3$, the predicted flux would be higher by a factor of ~ 5, while for $\gamma = 0.7$ it would be lower by a factor of ~ 3.

If it were not for astrophysical backgrounds, a telescope such as Fermi could definitively test an enormous range of dark matter models by observing the Galactic Center. The backgrounds from this direction of the sky, however, are quite bright, limiting the ultimate reach of such telescopes. The most important of these backgrounds consist of diffuse emission from cosmic-ray protons and electrons scattering with gas (via pion production and bremsstrahlung, respectively) and cosmic-ray electrons scattering with radiation (via inverse Compton scattering), along with emission from gamma-ray point sources, such as supernova remnants, pulsars, and the Milky Way's central supermassive black hole. Models for these backgrounds are constructed using inputs such as gas maps and the locally measured cosmic-ray spectrum. And while such models are often capable of describing the broad features of the observed galactic diffuse emission, they cannot (and should not be expected to) account for the detailed spectral or morphological characteristics of this background.

Fermi's observations of the Galactic Center have been used to place relatively stringent constraints on the dark matter's annihilation cross section [530, 531], comparable to those obtained from observations of dwarf galaxies (see below). More interestingly, these observations have also revealed an excess of gamma rays relative to expected backgrounds [532, 533, 534, 535, 536, 529, 530]. The spectrum and angular distribution of this signal are both consistent with the predictions from annihilating dark matter, with $m_X \sim 40 - 70$ GeV, $\sigma\, v \sim 10^{-26}\,\mathrm{cm}^3/\mathrm{s}$, and $\gamma \sim 1.1 - 1.3$ (for the case of annihilations to $b\bar{b}$). As of the time of writing, the origin of this signal remains hotly debated.

19.3.2 Dwarf galaxies

The Milky Way's dark matter halo contains a large number of smaller subhalos, the largest of which contain stars and constitute satellite galaxies. This satellite population of the Milky Way includes the classical dwarfs (Draco, Ursa Minor, Sculptor, Fornax, etc.), as well as several dozen ultra-faint galaxies that were discovered using data from modern surveys, including the Sloan Digital Sky Survey (SDSS) and the Dark Energy Survey (DES) [537, 538, 539].

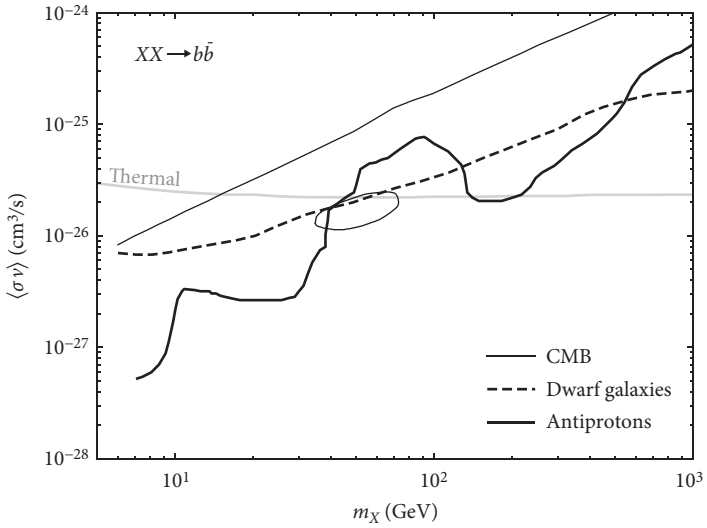


Figure 19.2 Upper limits on the dark matter's annihilation cross section to $b\bar{b}$ from observations of the CMB [8], gamma rays from dwarf galaxies [527], and cosmic-ray antiprotons [528]. The grey line is the cross section of a thermal relic and the closed region is the parameter space in which annihilating dark matter could be responsible for the Galactic Center gamma-ray excess [529].

Spectroscopic observations of dwarf galaxies can be used to measure the line of sight velocities of the brightest stars in these systems, providing us with information about the underlying dark matter distribution, and enabling us to estimate the annihilation J-factors of individual dwarfs [540, 541]. Such measurements yield J-factors in the range of $\sim 10^{16.2}$ to $10^{19.5}\,\mathrm{GeV}^2/\mathrm{cm}^5$ for most dwarfs, integrated over a circle of radius of 0.5° [540, 542]. Even for the most promising dwarfs, these J-factors and their corresponding gamma-ray fluxes are much smaller than those associated with the Galactic Center. Taking $J = 10^{19.5}\,\mathrm{GeV}^2/\mathrm{cm}^5$, for example, we obtain a flux of

$$F_\gamma \approx 2.7 \times 10^{-11}\,\mathrm{cm}^{-2}\mathrm{s}^{-1} \times \left(\frac{\langle \sigma\, v \rangle}{2.2 \times 10^{-26}\,\mathrm{cm}^3/\mathrm{s}}\right)\left(\frac{\int \frac{dN_\gamma}{dE_\gamma}\big|_{\mathrm{ann}} dE_\gamma}{10}\right)\left(\frac{100\,\mathrm{GeV}}{m_X}\right)^2.$$

This corresponds to a rate of approximately ~ 1.5 events per year for a telescope like Fermi, which is more than three orders of magnitude smaller than predicted from the direction of the Galactic Center.

Despite the relatively small gamma-ray fluxes that are expected from dark matter annihilating in dwarf galaxies, the low background rates from these objects make it possible for their observations to yield competitive constraints, especially after combining data from many different dwarf galaxies. The constraint derived from Fermi's observation of 28 dwarf galaxies [527] is shown in Figure 19.2, along with the region of parameter space that is favored to explain the Galactic Center gamma-ray excess [529].

19.4 Cosmic rays from dark matter annihilation

Most astrophysical sources of high-energy particles accelerate far more matter than antimatter, and this is reflected in the composition of the cosmic-ray spectrum. It has long

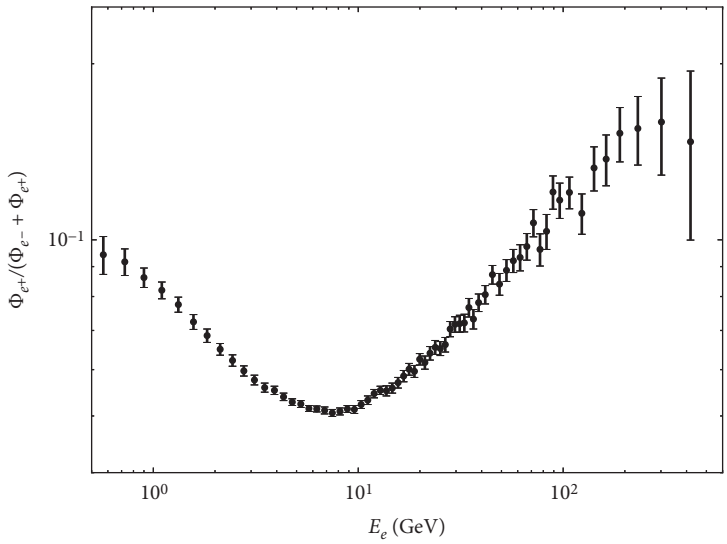


Figure 19.3 The cosmic-ray positron fraction, as measured by AMS [551].

been appreciated, however, that if dark matter particles are annihilating or decaying in the halo of the Milky Way, such processes would (in most models) produce equal amounts of matter and antimatter, potentially leading to an excess of positrons, antiprotons, or antinuclei relative to that predicted from standard astrophysical sources or mechanisms [543, 544, 545, 546, 547, 548].

Once cosmic rays are injected into the halo, they undergo diffusion, random walking through the Milky Way's tangled magnetic field, while also experiencing interactions that can cause them to lose energy, fragment, or decay. These processes can be described by the diffusion-loss equation, as given in Eq. 15.3. In addition to any particles that might be produced through dark matter annihilations, the source term in this equation, $Q(\vec{x}, E)$, includes contributions from astrophysical sources of cosmic rays (supernova remnants, pulsars, etc.), and from secondary production mechanisms, which account for those particles that are created through the interactions of other cosmic rays during their propagation through the interstellar medium. Secondary electrons and positrons, for example, are generated in the decays of pions that are produced in the collisions of hadronic cosmic rays with gas. The flux of these secondary cosmic rays as observed at earth can be calculated from Eq. 15.3 by setting $Q = \int dE_p \, J_p \, n_{\rm gas} \, d\sigma/dE$, where J_p is the flux of hadronic cosmic rays, $n_{\rm gas}$ is the gas density, and $d\sigma/dE$ is the differential cross section for the production of the secondary particles [549, 550].

19.4.1 Cosmic-ray positrons

The cosmic-ray positron spectrum has been measured in some detail by the space-based detectors AMS [551] and PAMELA [552]. To the surprise of many, these experiments have observed a steep rise in the ratio of cosmic-ray positrons to electrons-plus-positrons at energies between 10 and several hundred GeV (see Figure 19.3). In contrast, the fraction of these particles that consist of secondary positrons is predicted to decrease with energy. This observed rise in the cosmic-ray positron fraction thus indicates that sources of high-energy positrons must be present within the local volume of the Milky Way.

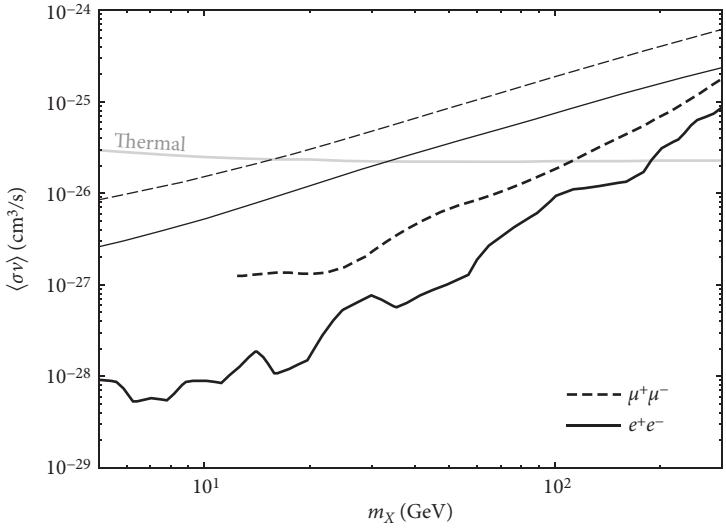


Figure 19.4 Upper limits on the dark matter's annihilation cross section to e^+e^- or $\mu^+\mu^-$ from measurements of the cosmic-ray positron spectrum [558, 559] (lower, thick) and the CMB [8] (upper, thin). The grey line is the annihilation cross section of a thermal relic.

When the rise in the cosmic-ray positron fraction was first reported, the possibility that this signal might be generated by annihilating dark matter received a great deal of attention. The current combination of cosmic-ray positron and electron data, however, can only be accommodated by dark matter that is quite heavy ($m_X \sim 1$–3 TeV) and that annihilates at a rate that would be in significant tension with modern gamma-ray data. A more likely explanation is that the bulk of these positrons are produced by nearby pulsars [553, 554]. In particular, recent observations by the HAWC Collaboration [555, 556] indicate that pulsars accelerate significant fluxes of very high-energy electrons and positrons, providing a consistent explanation for the cosmic-ray positron fraction as measured by AMS and PAMELA [557, 285].

Even if dark matter is not responsible for these observed positrons, it is possible to use measurements such as these to place constraints on its annihilation cross section, in particular in the case of dark matter particles that annihilate mostly to charged leptons. These constraints are shown in Figure 19.4 for dark matter that annihilates to e^+e^- or $\mu^+\mu^-$ [558, 559]. For masses up to several hundred GeV, these constraints are more stringent that those derived from measurements of the CMB.

19.4.2 Cosmic-ray antiprotons and antinuclei

In a small fraction of dark matter annihilations, antinucleons—or even antinuclei—could be produced. Unlike positrons, these particles experience negligible energy losses, and thus can reach us from much greater distances.

In addition to positrons, the AMS experiment has also produced a high-precision measurement of the cosmic-ray antiproton spectrum [560]. As only a small fraction of cosmic-ray collisions (or dark matter annihilations) are expected to produce an antinucleon, the ratio of antiprotons-to-protons in the cosmic-ray spectrum is quite small, typically

on the order of $\sim 10^{-4}$ [561]. Analyses of the measured antiproton-to-proton ratio, in conjunction with other secondary-to-primary ratios, have broadly found good agreement with the expectations from standard secondary production. These results have been used to place constraints on the dark matter annihilation cross section, in particular to quarks or gauge bosons [528, 562, 563]. These constraints are shown in Figure 19.2 for the case of annihilations to $b\bar{b}$ [528]. Qualitatively similar constraints are found for dark matter particles that annihilate to other hadronic channels, or to gauge or Higgs bosons.

In an even smaller fraction of dark matter annihilations, an anti-deuteron or an anti-helium nucleus could be produced. Although the fluxes of cosmic-ray antinuclei generated by dark matter would undoubtedly be very small, searches for anti-deuterons and anti-helium are expected to be almost background-free at low energies, providing a potentially powerful probe of annihilating dark matter [564]. The reason for this is that, while anti-deuterons and anti-helium can be produced in cosmic-ray collisions with gas, such interactions invariably take place in highly boosted frames of reference, leading to highly energetic outgoing particles. Dark matter annihilations, in contrast, take place at low velocities, potentially leading to the production of antinuclei without large amounts of kinetic energy.

19.5 Decaying dark matter

So far in this chapter, we have focused on annihilating dark matter and searches for the particles that might be produced in those annihilations. This choice is well motivated by the relationship between the abundance of a thermal relic and its annihilation cross section. But despite such arguments, there are many examples of viable dark matter candidates that do not appreciably annihilate. Alternatively, the particles that make up the dark matter could be unstable, leading to potentially observable decay products.

Observations of the CMB and large-scale structure indicate that the (comoving) abundance of dark matter has not appreciably changed over the course of our universe's history, at least since the beginning of the matter-dominated era. Such measurements have been used to constrain the lifetime of the particles that make up the dark matter, requiring $\tau_X \gtrsim 5 \times 10^{18}$ s, even if the decay products are invisible [395]. Much stronger constraints can be placed on dark matter candidates that produce detectable particles in their decays.

Unlike with the dark matter annihilation cross section, there is no clear benchmark target for the lifetime of a long-lived but unstable dark matter candidate. That being said, it could be argued that some values of this lifetime are more well motivated than others. For example, the lifetime of a particle that decays through a dimension-5 operator suppressed by the GUT scale ($M_{\rm GUT} \sim 10^{16}$ GeV) can be estimated as follows:

$$\tau \sim \frac{M_{\rm GUT}^2}{m_X^3} \sim 10^{20}\,{\rm s} \times \left(\frac{100\,{\rm keV}}{m_{\rm DM}}\right)^3. \tag{19.9}$$

From this we learn that dimension-5 operators, even if suppressed by a very high scale, tend to cause dark matter particles to decay on timescales that are already ruled out by cosmological considerations, unless they are quite light. On the other hand, if we consider

a decay that results from a dimension-6 operator, we arrive at the following estimate:

$$\tau \sim \frac{M_{\rm GUT}^4}{m_X^5} \sim 10^{25}\,{\rm s} \times \left(\frac{\rm TeV}{m_{\rm DM}}\right)^5. \tag{19.10}$$

This lifetime is not excluded on cosmological grounds, but could potentially be tested by searching for the dark matter's decay products. Observations of gamma-rays, X-rays, neutrinos, and cosmic rays have each been used to place constraints on the dark matter's lifetime [565].

To calculate the flux of gamma-rays from decaying dark matter, we modify Eq. 19.5, replacing the annihilation rate per volume, $\langle \sigma\, v \rangle \rho_X^2 / 2 m_X^2$, with the decay rate per volume, $\rho_X / m_X \tau_X$, and substituting the gamma-ray spectrum produced per annihilation with that produced per decay:

$$\frac{dN_\gamma}{dE_\gamma} = \frac{1}{4\pi\, \tau_X m_X} \left.\frac{dN_\gamma}{dE_\gamma}\right|_{\rm decay} \int_{\Delta\Omega} \int_{los} \rho_X(l, \Omega)\, dl\, d\Omega. \tag{19.11}$$

Because this flux is proportional to only one power of the dark matter density (as opposed to two powers in the case of dark matter annihilation), the best strategy is generally to study large regions of the sky, searching for decay products from throughout the halo of the Milky Way, and throughout the integrated volume of the observable universe. Due to our universe's opacity to gamma rays above ~ 1 TeV, the constraints on the dark matter's lifetime are approximately the same for all masses greater than $m_X \sim \mathcal{O}({\rm GeV})$, ruling out dark matter candidates that decay into Standard Model particles (other than neutrinos) with a lifetime less than $\sim 10^{28}$ s [566, 567, 568]

Additional reading

For a more detailed discussion of indirect searches for dark matter, I recommend Slatyer's lectures from the TASI summer school [569], as well as my own TASI lectures on the same topic [570].

Problems

19.1 Consider a dark matter candidate that is a cold thermal relic with an annihilation cross section that is p-wave suppressed, $\langle \sigma\, v \rangle \sim 2.2 \times 10^{-26}\,{\rm cm}^3/{\rm s} \times (v^2/0.1)$. If this particle undergoes kinetic decoupling at a temperature of $T_{\rm kd} = 10\,{\rm keV}$, estimate its velocity at the time of recombination. For what range of masses is such a particle excluded by CMB constraints (using $f_{\rm eff} = 0.5$)?

19.2 Derive the result given in Eq. 19.6.

19.3 Write a code to numerically integrate Eq. 19.5 over a line of sight and solid angle. For a WIMP that produces 10 photons per annihilation ($\int dN_\gamma/dE_\gamma\, dE_\gamma = 10$), calculate the number of photons per area and time from the innermost 1 degree around

the Galactic Center. Calculate the results using a generalized NFW profile with either $\gamma = 0.8$ or $\gamma = 1.4$ (using $R_s = 20\,\mathrm{kpc}$, and $\rho_X = 0.4\,\mathrm{GeV/cm^3}$ evaluated at $r = 8.25\,\mathrm{kpc}$).

19.4 Repeat Problem 19.3, but for the case of a halo profile that has a constant density core within 1 kpc of the Galactic Center (and that is described by an NFW profile elsewhere). How do your results compare to those found using a standard NFW profile?

19.5 Estimate the J-factor from the 1 degree radius circle centered around the the Andromeda galaxy. Other than its distance ($d = 773\,\mathrm{kpc}$), take the dark matter halo profile of Andromeda to be the same as that of the Milky Way.

19.6 Estimate the amount of time it would take for a high-energy positron to diffuse from the center of the Milky Way to our solar system. In comparing this to the timescale for inverse Compton and synchrotron energy losses, estimate the maximum energy of the positrons that can reach us from this distance.

19.7 Consider a 1 TeV dark matter particle that annihilates to e^+e^- with a cross section of $\langle \sigma v \rangle = 2.2 \times 10^{-26}\,\mathrm{cm^3/s}$. Solve the diffusion-loss equation (Eq. 15.3) to estimate the contribution from these annihilations to the local cosmic-ray positron flux. Treat the dark matter density in the local galaxy as uniform and neglect the Klein-Nishina suppression in the inverse Compton scattering cross section.

19.8 Consider a dark matter candidate that decays invisibly through a TeV-scale, dimension-5 operator. What is the maximum mass of this particle species? Could this particle be either a boson or a fermion?

20

Collider Physics for Astronomers

Much of what we have been able to infer about our universe's first second is based on measurements made at particle colliders. Many astronomers and astrophysicists, however, are never taught much about particle colliders, how they work, or what they are able to teach us about our universe. This chapter is designed specifically for this audience, and is intended to provide readers who are not experts in particle physics with a basic primer on the Large Hadron Collider and other particle accelerators, with an emphasis on their importance to cosmology.

20.1 Why collide particles?

Since the days of Ernest Rutherford, scientist have been studying the laws of physics by smashing particles into one another. The collisions of highly energetic particles are particularly useful in this respect. One reason for this is that the Heisenberg uncertainty principle limits the resolution that can be achieved in such collisions, $\Delta x \sim \hbar/2\Delta p$. To probe the structure of a nucleus with a radius of $r_N \sim (10^{-13}\,\text{cm}) \times A^{1/3}$, for example, requires particles with momenta that are not much less than $p \sim \hbar/2r_N \sim (10^2\,\text{MeV}) \times A^{-1/3}$.

A second reason why it is useful to collide high-energy particles is that, if enough energy is present in a collision, it may be possible for new particles to be created. Consider, for example, a beam of highly relativistic particles, each of energy E_{beam}, striking some stationary particles of mass m_{target}. In the center-of-momentum frame, the total energy of these collisions is given by:

$$\begin{aligned} E_{\text{cm}} &= \sqrt{E_{\text{tot}}^2 - p_{\text{tot}}^2} \\ &= \sqrt{(\gamma_{\text{beam}} m_{\text{beam}} + m_{\text{target}})^2 - (\gamma_{\text{beam}} m_{\text{beam}} v_{\text{beam}})^2} \\ &= \sqrt{m_{\text{beam}}^2 + m_{\text{target}}^2 + 2E_{\text{beam}} m_{\text{target}}} \\ &\approx \sqrt{2E_{\text{beam}} m_{\text{target}}}, \end{aligned} \tag{20.1}$$

where $E_{\rm tot}$ is the total energy of the colliding particles in the lab frame, and in the last step I have assumed that $E_{\rm beam} \gg m_{\rm target}$. To produce massive particles in such collisions, the total energy in the center-of-momentum frame must exceed the mass of the newly created particles, requiring a large product of beam energy and target mass. This kind of setup, in which a particle beam is directed onto a stationary target, is known as a fixed target experiment.

Since the center-of-momentum energy of a fixed target collision grows only with the square root of the beam energy, $E_{\rm cm} \propto E_{\rm beam}^{1/2}$, such experiments are generally limited in their ability to produce massive particles. One can do better in this respect by directing two beams of relativistic particles head-on into one another. In this case, the total energy of a collision in the center-of-momentum frame is given as follows:

$$\begin{aligned} E_{\rm cm} &= \sqrt{E_{\rm tot}^2 - p_{\rm tot}^2} \\ &= \sqrt{(\gamma_1 m_1 + \gamma_2 m_2)^2 - (\gamma_1 m_1 v_1 - \gamma_2 m_2 v_2)^2} \\ &= \sqrt{m_1^2 + m_2^2 + 2E_1 E_2 + 2p_1 p_2} \\ &\approx 2\sqrt{E_1 E_2}. \end{aligned} \tag{20.2}$$

For two beams of equal energy, for example, the total energy in the center-of-momentum frame is simply equal to twice the beam energy. At the Large Hadron Collider (LHC), protons are collided with $E_1 = E_2 \approx 6.8$ TeV, for a total energy of $E_{\rm cm} \approx 13.6$ TeV. In contrast, if one of the LHC's beams were to be directed at a target of stationary protons, the resulting collisions would achieve a much more modest energy of only $E_{\rm cm} \approx 113$ GeV.

Fixed target and collider experiments each have their respective advantages and disadvantages. In particular, despite their limited ability to achieve high center-of-momentum energies (and thus to produce massive particles), fixed target experiments can be used to study far more collisions than can be observed in collider environments. So, while colliders are generally best suited for producing heavy particles and otherwise studying the laws of physics at high energies, fixed target experiments are often able to observe and study processes that occur less frequently, such as those that involve feeble couplings.

20.2 Electron and proton colliders

Particle accelerators use electromagnetic fields to propel charged particles to high speeds and to maintain those particles in tightly focused beams. Many accelerators circulate the particle beams around a ring, making it possible to recycle the beam and to attempt to collide the particles within it many times. Accelerating a charged particle around a ring, however, also causes those particles to lose energy through the process of synchrotron radiation.

To maintain a beam in a circular path of radius R, the transverse force from the magnetic field must balance against the centrifugal force:

$$-evB = -\gamma m \frac{v^2}{R}, \tag{20.3}$$

where e and m are the charge and mass of the accelerated particles, v and γ are the velocity and Lorentz factor of the beam, and B is the strength of the transverse magnetic field. To

maintain a beam of particles at a given energy thus requires a magnetic field of strength $B = \gamma m v/eR = Ev/eR$.

In the presence of this magnetic field, charged particles lose energy to synchrotron emission at the following rate (see Sec. 5.1.2):

$$\begin{aligned}\frac{dE}{dt} &= \frac{2E^2e^4B^2}{3m^4} \\ &= \frac{2E^4e^2v^2}{3m^4R^2} \\ &\approx 27\,\text{TeV/s} \times \left(\frac{E}{100\,\text{GeV}}\right)^4 \left(\frac{e}{1}\right)^2 \left(\frac{m_e}{m}\right)^4 \left(\frac{26.7\,\text{km}}{2\pi R}\right)^2.\end{aligned} \tag{20.4}$$

Note that I have scaled this result to the size of the LHC's tunnel, $2\pi R \approx 26.7$ km. In one revolution, $\Delta t \approx 2\pi R/c$, a particle in the beam will lose the following fraction of its energy:

$$\begin{aligned}\frac{\Delta E}{E} &= \frac{4\pi E^3e^2v^2}{3m^4R} \\ &\approx 0.024 \times \left(\frac{E}{100\,\text{GeV}}\right)^3 \left(\frac{e}{1}\right)^2 \left(\frac{m_e}{m}\right)^4 \left(\frac{26.7\,\text{km}}{2\pi R}\right).\end{aligned} \tag{20.5}$$

From this, we see that electrons with $E_e \gtrsim 100\,\text{GeV}$ will lose their energy fairly rapidly in a ring of this size. This ultimately limits the energy to which electrons and positrons can be accelerated in a circular collider. To date, the highest energy electron-positron collider has been Large Electron-Positron (LEP) collider at CERN. LEP was located in the same circular tunnel that the LHC now occupies, and achieved a maximum center-of-momentum energy of 209 GeV. To reach higher energies, a linear collider could be used, although this would have the disadvantage of only being able to use the beam once. Alternatively, one could achieve greater energies by accelerating and colliding particles with more mass, such as protons or nuclei (or perhaps, in the future, muons).

The main advantage of using electrons and positrons in this context is that their collisions are relatively clean and straightforward to study. While operating at its maximum energy, every collision at LEP took place between an electron and a positron, each with an energy of 104.5 GeV. By adjusting the energy of the LEP beams, it was possible to study the energy dependence of the electron-positron cross section to various final states, and to look for any thresholds or resonances that might be present. In contrast, the collisions at a proton collider do not take place between the entire protons, but between individual partons—gluons, quarks, and antiquarks[1]—each of which carries only a fraction of its proton's total energy. As a result, the total energy of a colliding pair of partons at the LHC varies considerably from collision to collision, while always being less than 13.6 TeV. Furthermore, the complexities of QCD often cause collisions involving protons to be more messy than electron-positron collisions, featuring individual events with large numbers of

[1] A given nucleon contains three valence quarks, but also many virtual gluons and quark-antiquark pairs, known as sea quarks. The measured distributions of these quarks and gluons are quantified in terms of what are known as parton distribution functions.

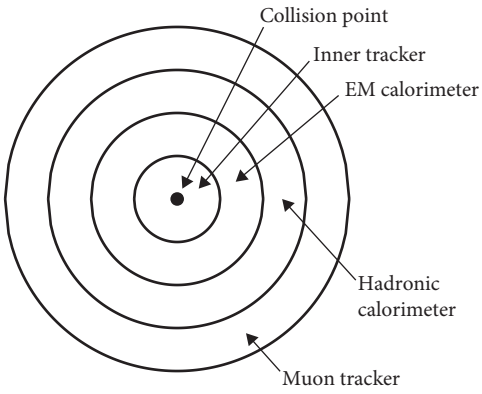


Figure 20.1 A schematic depiction of one of the LHC's particle detectors, as seen looking along the axis of the proton beam.

quark and gluon jets.[2] For these reasons, measurements made at proton colliders are often less precise than those performed at lepton colliders.

20.3 Particle detection at the LHC

In a collision at the LHC, the interacting protons exchange a significant amount of momentum and subsequently explode as a shower of energetic particles. By measuring the characteristics of these outgoing particles, one can reconstruct the dynamics of the underlying collision, ideally allowing us to infer the characteristics of the particles that were created in the initial scattering event.

Each of the collision points at the LHC is surrounded by a large particle detector, designed to measure the energy, momentum, and electric charge of the outgoing particles. These include the general-purpose ATLAS and CMS detectors, as well as the more specialized experiments known as ALICE and LHCb. While each of these detectors is unique, they all consist of a similar combination of basic components that provide us with different information about the particles that pass through the experiment.

Immediately surrounding each collision point is a tracker, which uses the effects of ionization to measure and record the trajectories of electrically charged particles. By embedding the tracker in a magnetic field, it is possible to use a particle's trajectory to infer its charge and momentum.

Beyond the tracker, particles encounter regions of the detector known as the electromagnetic and hadronic calorimeters. These components detect and measure the energy that is deposited in the form of electromagnetic and hadronic radiation, respectively. More specifically, the electromagnetic calorimeter stops electrons, positrons, and photons, primarily through the effects of bremsstrahlung, measuring how much energy is deposited in this way. The hadronic calorimeter functions in a similar fashion, stopping mesons and baryons through their inelastic collisions with atomic nuclei. While hadrons are able to pass through the electromagnetic calorimeter without losing much of their energy, the higher

[2] A jet is a narrow cone of particles that is produced through the showering and fragmentation of an energetic quark or gluon.

density hadronic calorimeter rapidly slows these particles, bringing them to a near halt and measuring the energy that is deposited in this process.

The outermost component of the LHC's detectors is a muon tracker. Being much more massive than electrons, muons lose relatively little of their energy through ionization or bremsstrahlung, and thus penetrate the inner tracker and calorimeters without losing much of their initial energy. Once a muon leaves the calorimeters, however, it enters a second higher-density tracker, which is again embedded in a magnetic field. Much as the inner tracker does for electrons, this system is designed to measure the momentum and charge of the muons that pass through it.

These detector components are collectively able to detect and characterize many of the photons, electrons, muons, and hadrons that emerge from the proton-proton collisions at the LHC. But what about the other Standard Model particles? Any top quarks or $W^{\pm}$, Z, or Higgs bosons that are produced at the LHC will decay very quickly, well before they enter any of the detector components. Even tau leptons decay before reaching the inner tracker. Furthermore, any light quarks or gluons that are created in these collisions quickly hadronize, forming a collection of hadrons and other stable or long-lived particles. The particles that enter the detector thus consist almost entirely of electrons, photons, muons, and hadrons, along with neutrinos that pass through the detector without interacting.

Although short-lived particles cannot be directly observed at the LHC, their presence can often be inferred from the kinematics of their decay products. Consider, for example, a Z boson that decays into a pair of electrons or muons. By measuring the momenta of these outgoing leptons, one can determine the invariant mass of the lepton pair, which, in turn, can be used to infer the mass of the particle that produced them in their decay.

While neutrinos escape the experiment without being detected, even their presence can be inferred from the kinematics of a given event. In particular, since the colliding protons have nonzero momenta only along the axis of the beam, the total transverse momentum of the outgoing particles must be equal to zero. If the observed particles are instead measured to have a finite total transverse momentum, this would imply that some momentum must have been carried away by undetected particles, such as neutrinos.

20.4 Cross sections and event rates at the LHC

The LHC collides protons together with a center-of-momentum energy of 13.6 TeV. At this energy, the total cross section for a proton-proton collision is approximately $\sigma_{pp} \approx 1.1 \times 10^{-25}\,\text{cm}^2$. In collider physics, it is common to describe cross sections in units of barns, where $1\,\text{b} \equiv 10^{-24}\,\text{cm}^2$.[3] The total cross section at the LHC is therefore approximately $\sigma_{pp} \approx 110\,\text{mb}$.

The rate at which collisions occur at the LHC is given by the total cross section multiplied by a quantity known as the "luminosity." The LHC is designed to operate at a luminosity of $L = 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ (and at $L = 10^{35}\,\text{cm}^{-2}\text{s}^{-1}$ during its high-luminosity run, scheduled to begin in 2027). This means that at its current design luminosity, approximately $\sigma_{pp} \times L \approx 1.1 \times 10^9$ collisions take place at the LHC each second, for a total of 3.5×10^{16} collisions in a full year of operation.

[3] The barn as a unit of area was coined by Enrico Fermi during the Manhattan Project, for which the largest relevant nuclear cross sections were roughly $\sigma \sim 1\,\text{b}$. The name of this unit was intended to suggest that colliding a particle with a large nucleus was as easy as "hitting the broad side of a barn."

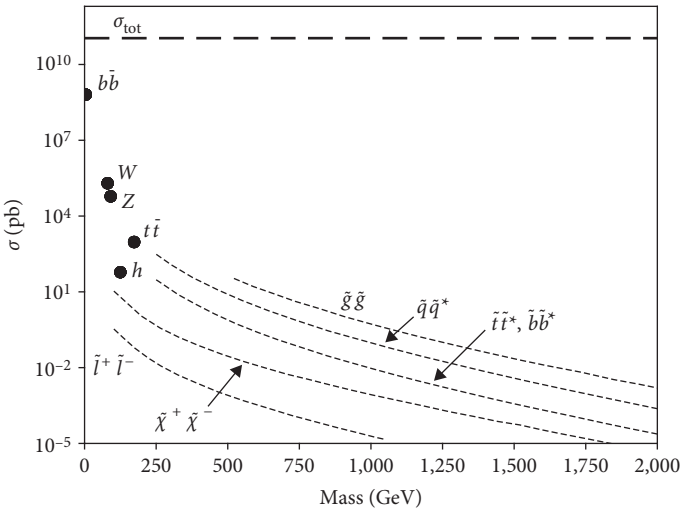


Figure 20.2 The production cross sections for various Standard Model particles, in 13.6 TeV proton-proton collisions. The dashed curves represent the approximate cross sections for the production of various species of supersymmetric particles, as a function of those particles' mass.

While 3.5×10^{16} collisions per year might seem like an enormous number, this rate is actually a limiting factor for many of the searches for new physics that we might want to perform with this machine. In most of the LHC's collisions, the scattering actually takes place between two relatively low-energy gluons, leading to a messy event that does little to advance our understanding of the laws of physics. The most useful and interesting events, which are far more rare, involve a collision between a pair of high-energy partons, potentially allowing for the production of massive particles.

As an example, consider the Higgs boson, which was discovered at the LHC in 2012 [571, 572]. The cross section for the production of a Higgs boson in a 13.6 TeV proton-proton collision is about 60 pb. This means that a Higgs boson is produced in only one out of every two billion collisions at the LHC. Once such a particle is created, it decays almost instantly, usually into pairs of bottom quarks or gauge bosons. The cleanest signal of a Higgs boson, however, arises in those relatively rare instances in which the Higgs decays into a pair of photons—such events can be distinguished from backgrounds by exploiting the fact that the invariant mass of the photon pair is approximately equal to the mass of the Higgs boson. Unfortunately, the branching fraction for $h \to \gamma\gamma$ is only about 2×10^{-3}. Putting this together, we estimate that $\sim 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1} \times 60\,\mathrm{pb} \approx 1.9 \times 10^{7}$ Higgs bosons are produced at the LHC each year, of which $\sim (1.9 \times 10^{7}) \times (2 \times 10^{-3}) \approx 4 \times 10^{4}$ decay through the $h \to \gamma\gamma$ channel. After accounting for experimental efficiencies (which are typically at a level of $\sim 10^{-2}$), we are left with only a few hundred of these events to study. While this rate is, in fact, high enough to facilitate the discovery of the Higgs boson at the LHC, this exercise helps to illustrate why such large numbers of collisions are needed in order for this machine to be sensitive to very massive or very feebly interacting particles.

Compared to the Higgs boson, other Standard Model particles are all produced at significantly higher rates at the LHC (see Figure 20.2). This is due to factors including the large mass and small couplings of the Higgs boson, as well as to the fact that this particle is produced at the LHC mostly through a diagram involving a top quark loop. Even top quarks,

which are the most massive particles in the Standard Model, are produced at a rate that is more than an order of magnitude higher than that of the Higgs boson.

All of the particle species described by the Standard Model are produced in large numbers at the LHC and, in many respects, this machine provides us with our most powerful way of probing the properties of the Standard Model. But what about particles beyond the boundaries of the Standard Model? Which kinds of exotic particles should we expect the LHC to be sensitive to? The answer to this question depends on many factors, including the masses and couplings of the particles in question. For particles with QCD color (such as squarks or gluinos in the context of supersymmetry, for example), the production cross sections are generally expected to be quite large (see Figure 20.2), allowing the LHC to rule out the existence of any such particles lighter than $\sim$ 1–2 TeV (the precise value of this constraint depends on the spin, degrees of freedom, and decay channels of the particles in question). Constraints on electrically charged particles are significantly weaker, generally excluding masses up to a few hundred GeV. Particles that are only weakly interacting are much less strongly constrained by the LHC and other collider-based experiments.

20.5 The LHC and the early universe

Measurements of the light element abundances allow us to constrain the state of our universe during the era of BBN. More specifically, this data provides us with a measurement of our universe's expansion rate at the time of proton-neutron freeze out, around a second after the Big Bang, confirming that our universe was dominated at that time by a bath of relativistic Standard Model particles, with a temperature of $T \sim 1$ MeV.

Prior to BBN, we have essentially no way of directly determining what forms of matter and energy were present in our universe. The best we can do is to take what we have learned about the laws of physics from the LHC and other laboratory experiments, and use this information to infer as much as we can about the state of our universe during its first fraction of a second. Unfortunately, the information provided by such experiments is in many respects incomplete, leaving us with many open questions about the first moments of our universe's history.

Consider, for example, our universe at a time $\sim 10^{-10}$ s after the Big Bang. If we assume that the energy density of our universe at that time was dominated by the forms of matter and energy that we have observed and studied at colliders, we can use the first Friedmann equation, along with thermodynamic considerations, to deduce that a $T \sim 100$ GeV thermal bath should have been present at this time, containing equilibrium abundances of all of the particle species described by the Standard Model.

How much could our universe at $t \sim 10^{-10}$ s have departed from this picture? From the null results of searches for new physics at the LHC, we can be confident that no exotic particles with electric charge or QCD color were being produced in large numbers at this time—the constrains on such particles extend to masses well above $T \sim 100$ GeV. But what about more feebly interacting particles? At $T \sim 100$ GeV, a particle species would reach equilibrium with the thermal bath even if its couplings to the Standard Model were *very* small. Up to order 1 factors, the condition for equilibrium, $\sigma n v \gtrsim H$, can be written as

$$\begin{aligned} \sigma &\gtrsim \left(\frac{8\pi^3}{3g_\star(T)}\right)^{1/2} \frac{1}{M_{\rm Pl} T} \\ &\gtrsim 10^{-13}\,{\rm pb} \times \left(\frac{g_\star(T)}{100}\right)^{1/2} \left(\frac{100\,{\rm GeV}}{T}\right). \end{aligned} \tag{20.6}$$

This cross section is absolutely tiny, demonstrating that a wide range of feebly interacting particles could have been produced in large numbers in the early universe, without leading to any detectable signals at the LHC. If a large number of such particle species exist (perhaps constituting a hidden sector, without any Standard Model gauge charges), these forms of matter and energy could even have dominated the energy density of the early universe.

Another interesting possibility involves particles that are too heavy to be produced at the LHC. If very feebly interacting, such particles could have been produced at very early times, only to later decouple from the thermal bath as our universe expanded and cooled. If sufficiently long-lived, these particles could have become nonrelativistic before decaying, potentially coming to dominate the energy density of the early universe. If such particles are stable, they could be the dark matter. If not, their decays could have reheated the early universe, or even populated it with other exotic forms of matter or energy (see Sec. 9.7.2).

Additional reading

For an accessible introduction to collider physics, I recommend Larkoski's *Elementary Particle Physics*, as well as the TASI lectures by Strassler [573] and Schwartz [574].

Problems

20.1 Show that the third lines of Eqs. 20.1 and 20.2 follow from the second lines of their corresponding expressions.

20.2 If a muon-antimuon collider were constructed in the LHC tunnel, at what maximum center-of-momentum energy would these particles lose 1% or less of their energy in each revolution? What kinds of challenges might operating a muon collider pose?

20.3 Imagine that, in a collision at the LHC, we observe an electron with a momentum of $\vec{p}_1 = (100\,\text{GeV}, 0, 0)$ and a positron with a momentum of $\vec{p}_2 = (10\,\text{GeV}, 50.6\,\text{GeV}, 0)$, where the beam is aligned along the z-axis. What can we deduce about the origin of these particles? Hint: Calculate the invariant mass of the electron-positron pair.

20.4 Consider a pair of b quarks that is produced at the LHC, each with an energy of 100 GeV. How long do we expect such particles to travel from the collision point before they decay? Given that the inner tracker begins a few centimeters from the collision point, how could the finite lifetime of the b quark be used to distinguish the decay products of these particles from other kinds of jets?

20.5 The main way that Higgs bosons are produced at the LHC is through the process of gluon-gluon fusion, in which a pair of gluons create a Higgs boson through a top quark loop. Make an order of magnitude estimate of the cross section for the process $gg \to h \to \gamma\gamma$. Compare this result to the cross section for Higgs production and the branching fraction for Higgs decay into a pair of photons mentioned in the text.

20.6 Making the approximation that the LHC could be used to discover any particle species that it can produce at a rate of $\sim 10^4$ or more per year, use the cross sections given in Figure 20.2 to estimate the largest mass of the gluinos, squarks, stops/sbottoms, charginos, and sleptons that could be discovered by this machine.

Bibliography

[1] H. Shapley and H. D. Curtis, "The scale of the universe," *Bulletin of the National Research Council* **2** no. 11, (May, 1921) 171–217.

[2] A. Einstein, "The field equations of gravitation," *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1915** (1915) 844–847.

[3] H. S. Leavitt and E. C. Pickering, "Periods of 25 variable stars in the small Magellanic cloud," *Harvard Obs. Circ.* **173** (1912) 1–3.

[4] V. M. Slipher, "Nebulae," *Proc. Am. Phil. Soc.* **56** (1917) 403–409.

[5] E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," *Proc. Nat. Acad. Sci.* **15** (1929) 168–173.

[6] F. Hoyle, "A new model for the expanding universe," *Mon. Not. Roy. Astron. Soc.* **108** (Jan., 1948) 372.

[7] H. Bondi and T. Gold, "The steady-state theory of the expanding universe," *Mon. Not. Roy. Astron. Soc.* **108** (Jan., 1948) 252.

[8] **Planck** Collaboration, N. Aghanim et al., "Planck 2018 results. VI. Cosmological parameters," *Astron. Astrophys.* **641** (2020) A6, arXiv:1807.06209 [astro-ph.CO].

[9] A. G. Riess et al., "A 2.4% Determination of the Local Value of the Hubble Constant," *Astrophys. J.* **826** no. 1, (2016) 56, arXiv:1604.01424 [astro-ph.CO].

[10] A. G. Riess, S. Casertano, W. Yuan, L. M. Macri, and D. Scolnic, "Large magellanic cloud cepheid standards provide a 1% foundation for the determination of the Hubble constant and stronger evidence for physics beyond ΛCDM," *Astrophys. J.* **876** no. 1, (2019) 85, arXiv:1903.07603 [astro-ph.CO].

[11] K. C. Wong et al., "H0LiCOW—XIII. A 2.4 per cent measurement of H0 from lensed quasars: 5.3σ tension between early- and late-universe probes," *Mon. Not. Roy. Astron. Soc.* **498** no. 1, (2020) 1420–1439, arXiv:1907.04869 [astro-ph.CO].

[12] A. Einstein, "Cosmological considerations in the general theory of relativity," *Sitzungsber. Preuss. Akad. Wiss. Berlin (Math. Phys.)* **1917** (1917) 142–152.

[13] D. Lovelock, "The Einstein tensor and its generalizations," *J. Math. Phys.* **12** (1971) 498–501.

[14] G. 't Hooft, "Symmetry breaking through Bell-Jackiw anomalies," *Phys. Rev. Lett.* **37** (1976) 8–11.

[15] G. 't Hooft, "Computation of the quantum effects due to a four-dimensional pseudoparticle," *Phys. Rev. D* **14** (1976) 3432–3450. [Erratum: *Phys. Rev. D* 18, 2199 (1978)].

[16] R. E. Shrock, "General theory of weak processes involving neutrinos. 2. pure leptonic decays," *Phys. Rev. D* **24** (1981) 1275.

[17] P. B. Pal and L. Wolfenstein, "Radiative decays of massive neutrinos," *Phys. Rev. D* **25** (1982) 766.

[18] D. J. Gross and F. Wilczek, "Ultraviolet behavior of nonabelian gauge theories," *Phys. Rev. Lett.* **30** (1973) 1343–1346.
[19] H. D. Politzer, "Reliable perturbative results for strong interactions?," *Phys. Rev. Lett.* **30** (1973) 1346–1349.
[20] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.* **19** (1967) 1264–1266.
[21] P. Gondolo and G. Gelmini, "Cosmic abundances of stable particles: Improved analysis," *Nucl. Phys. B* **360** (1991) 145–179.
[22] H. Bethe and W. Heitler, "On the stopping of fast particles and on the creation of positive electrons," *Proc. Roy. Soc. Lond. A* **146** (1934) 83–112.
[23] G. R. Blumenthal and R. J. Gould, "Bremsstrahlung, synchrotron radiation, and Compton scattering of high-energy electrons traversing dilute gases," *Rev. Mod. Phys.* **42** (1970) 237–270.
[24] **Particle Data Group** Collaboration, P. A. Zyla et al., "Review of particle physics," *PTEP* **2020** no. 8, (2020) 083C01.
[25] A. W. Strong, I. V. Moskalenko, and V. S. Ptuskin, "Cosmic-ray propagation and interactions in the galaxy," *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 285–327, arXiv:astro-ph/0701517.
[26] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, "A new version of the event generator Sibyll," *PoS* **ICRC2015** (2016) 558, arXiv:1510.00568 [hep-ph].
[27] S. Ostapchenko, "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model," *Phys. Rev. D* **83** (2011) 014018, arXiv:1010.1869 [hep-ph].
[28] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider," *Phys. Rev. C* **92** no. 3, (2015) 034906, arXiv:1306.0121 [hep-ph].
[29] G. Breit and J. A. Wheeler, "Collision of two light quanta," *Phys. Rev.* **46** no. 12, (1934) 1087–1091.
[30] F. A. Aharonian, A. M. Atoian, and A. M. Nagapetian, "Photoproduction of electron-positron pairs in compact X-ray sources," *Astrofizika* **19** (Apr., 1983) 323–334.
[31] R. J. Gould and G. P. Schreder, "Pair production in photon-photon collisions," *Phys. Rev.* **155** (1967) 1404–1407.
[32] S. Seager, D. D. Sasselov, and D. Scott, "How exactly did the universe become neutral?," *Astrophys. J. Suppl.* **128** (2000) 407–430, arXiv:astro-ph/9912182.
[33] D. J. Fixsen, E. S. Cheng, J. M. Gales, J. C. Mather, R. A. Shafer, and E. L. Wright, "The cosmic microwave background spectrum from the full COBE FIRAS data set," *Astrophys. J.* **473** (1996) 576, arXiv:astro-ph/9605054.
[34] D. J. Fixsen, "The temperature of the cosmic microwave background," *Astrophys. J.* **707** (2009) 916–920, arXiv:0911.1955 [astro-ph.CO].
[35] **Planck** Collaboration, N. Aghanim et al., "Planck 2018 results. V. CMB power spectra and likelihoods," *Astron. Astrophys.* **641** (2020) A5, arXiv:1907.12875 [astro-ph.CO].
[36] S. Dodelson, "Coherent phase argument for inflation," *AIP Conf. Proc.* **689** no. 1, (2003) 184–196, arXiv:hep-ph/0309057.
[37] J. Silk, "Cosmic black body radiation and galaxy formation," *Astrophys. J.* **151** (1968) 459–471.
[38] G. Efstathiou, "Cosmological perturbations," in *36th Scottish Universities Summer School in Physics, a NATO ASI: Physics of the Early Universe*. 1989.
[39] R. A. Sunyaev and Y. B. Zeldovich, "Microwave background radiation as a probe of the contemporary structure and history of the universe," *Ann. Rev. Astron. Astrophys.* **18** (1980) 537–560.
[40] A. Lewis and A. Challinor, "CAMB: Code for anisotropies in the microwave background." Astrophysics Source Code Library, record ASCL:1102.026, Feb., 2011.
[41] M. Kamionkowski, A. Kosowsky, and A. Stebbins, "A Probe of primordial gravity waves and vorticity," *Phys. Rev. Lett.* **78** (1997) 2058–2061, arXiv:astro-ph/9609132.
[42] A. Lue, L.-M. Wang, and M. Kamionkowski, "Cosmological signature of new parity violating interactions," *Phys. Rev. Lett.* **83** (1999) 1506–1509, arXiv:astro-ph/9812088.

[43] **SPT** Collaboration, A. T. Crites et al., "Measurements of e-mode polarization and temperature-e-mode correlation in the cosmic microwave background from 100 square degrees of SPTpol data," *Astrophys. J.* **805** no. 1, (2015) 36, arXiv:1411.1042 [astro-ph.CO].

[44] **ACTPol** Collaboration, S. Naess et al., "The Atacama Cosmology Telescope: CMB polarization at $200 < \ell < 9000$," *JCAP* **10** (2014) 007, arXiv:1405.5524 [astro-ph.CO].

[45] **ACTPol** Collaboration, T. Louis et al., "The Atacama Cosmology Telescope: Two-season ACTPol spectra and parameters," *JCAP* **06** (2017) 031, arXiv:1610.02360 [astro-ph.CO].

[46] **CMB-S4** Collaboration, K. N. Abazajian et al., "CMB-S4 science book, first edition," arXiv:1610.02743 [astro-ph.CO].

[47] F. Zwicky, "Die Rotverschiebung von extragalaktischen Nebeln," *Helv. Phys. Acta* **6** (1933) 110–127.

[48] F. Zwicky, "On the masses of nebulae and of clusters of nebulae," *Astrophys. J.* **86** (1937) 217–246.

[49] V. C. Rubin and W. K. Ford, Jr., "Rotation of the Andromeda nebula from a spectroscopic survey of emission regions," *Astrophys. J.* **159** (1970) 379–403.

[50] J. P. Ostriker and P. J. E. Peebles, "A numerical study of the stability of flattened galaxies: Or, can cold galaxies survive?," *Astrophys. J.* **186** (1973) 467–480.

[51] B. Paczynski, "Gravitational microlensing by the galactic halo," *Astrophys. J.* **304** (1986) 1–5.

[52] **MACHO** Collaboration, C. Alcock et al., "The MACHO project: Microlensing results from 5.7 years of LMC observations," *Astrophys. J.* **542** (2000) 281–307, arXiv:astro-ph/0001272.

[53] **EROS** Collaboration, T. Lasserre, "Not enough stellar mass machos in the galactic halo," *Astron. Astrophys.* **355** (2000) L39–L42, arXiv:astro-ph/0002253.

[54] **EROS-2** Collaboration, P. Tisserand et al., "Limits on the macho content of the galactic halo from the EROS-2 survey of the Magellanic clouds," *Astron. Astrophys.* **469** (2007) 387–404, arXiv:astro-ph/0607207.

[55] M. Milgrom, "A modification of the Newtonian dynamics as a possible alternative to the hidden mass hypothesis," *Astrophys. J.* **270** (1983) 365–370.

[56] M. Milgrom, "A modification of the Newtonian dynamics: Implications for galaxies," *Astrophys. J.* **270** (1983) 371–383.

[57] M. Milgrom, "A modification of the Newtonian dynamics: Implications for galaxy systems," *Astrophys. J.* **270** (1983) 384–389.

[58] J. D. Bekenstein, "Relativistic gravitation theory for the MOND paradigm," *Phys. Rev. D* **70** (2004) 083509, arXiv:astro-ph/0403694. [Erratum: *Phys. Rev. D* 71, 069901 (2005)].

[59] D. Clowe, M. Bradac, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A direct empirical proof of the existence of dark matter," *Astrophys. J. Lett.* **648** (2006) L109–L113, arXiv:astro-ph/0608407.

[60] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama, "Constraints on primordial black holes," arXiv:2002.12778 [astro-ph.CO].

[61] W. Hu, R. Barkana, and A. Gruzinov, "Cold and fuzzy dark matter," *Phys. Rev. Lett.* **85** (2000) 1158–1161, arXiv:astro-ph/0003365.

[62] L. Hui, J. P. Ostriker, S. Tremaine, and E. Witten, "Ultralight scalars as cosmological dark matter," *Phys. Rev. D* **95** no. 4, (2017) 043541, arXiv:1610.08297 [astro-ph.CO].

[63] K. Schutz, "Subhalo mass function and ultralight bosonic dark matter," *Phys. Rev. D* **101** no. 12, (2020) 123026, arXiv:2001.05503 [astro-ph.CO].

[64] N. Dalal and A. Kravtsov, "Excluding fuzzy dark matter with sizes and stellar kinematics of ultrafaint dwarf galaxies," *Phys. Rev. D* **106** no. 6, (2022) 063517, arXiv:2203.05750 [astro-ph.CO].

[65] S. Tremaine and J. E. Gunn, "Dynamical role of light neutral leptons in cosmology," *Phys. Rev. Lett.* **42** (1979) 407–410.

[66] C. Boehm, M. J. Dolan, and C. McCabe, "A lower bound on the mass of cold thermal dark matter from Planck," *JCAP* **08** (2013) 041,arXiv:1303.6270 [hep-ph].
[67] K. M. Nollett and G. Steigman, "BBN and the CMB constrain neutrino coupled light WIMPs," *Phys. Rev. D* **91** no. 8, (2015) 083505, arXiv:1411.6005 [astro-ph.CO].
[68] K. Griest and M. Kamionkowski, "Unitarity limits on the mass and radius of dark matter particles," *Phys. Rev. Lett.* **64** (1990) 615.
[69] K. Griest and D. Seckel, "Three exceptions in the calculation of relic abundances," *Phys. Rev. D* **43** (1991) 3191–3203.
[70] D. E. Kaplan, M. A. Luty, and K. M. Zurek, "Asymmetric dark matter," *Phys. Rev. D* **79** (2009) 115016, arXiv:0901.4117 [hep-ph].
[71] M. L. Graesser, I. M. Shoemaker, and L. Vecchi, "Asymmetric WIMP dark matter," *JHEP* **10** (2011) 110, arXiv:1103.2771 [hep-ph].
[72] K. Petraki and R. R. Volkas, "Review of asymmetric dark matter," *Int. J. Mod. Phys. A* **28** (2013) 1330028, arXiv:1305.4939 [hep-ph].
[73] L. J. Hall, K. Jedamzik, J. March-Russell, and S. M. West, "Freeze-in production of FIMP dark matter," *JHEP* **03** (2010) 080, arXiv:0911.1120 [hep-ph].
[74] T. Lin, "Dark matter models and direct detection," *PoS* **333** (2019) 009, arXiv:1904.07915 [hep-ph].
[75] G. Bertone and D. Hooper, "History of dark matter," *Rev. Mod. Phys.* **90** no. 4, (2018) 045002, arXiv:1605.04909 [astro-ph.CO].
[76] M. Escudero, D. Hooper, G. Krnjaic, and M. Pierre, "Cosmology with a very light L_μ-L_τ gauge boson," *JHEP* **03** (2019) 071, arXiv:1901.02010 [hep-ph].
[77] D. Hooper, G. Krnjaic, and S. D. McDermott, "Dark radiation and superheavy dark matter from black hole domination," *JHEP* **08** (2019) 001, arXiv:1905.01301 [hep-ph].
[78] N. Y. Gnedin and L. Hui, "Probing the universe with the Lyman alpha forest: 1. Hydrodynamics of the low density IGM," *Mon. Not. Roy. Astron. Soc.* **296** (1998) 44–55, arXiv:astro-ph/9706219.
[79] S. Naoz and R. Barkana, "The formation and gas content of high redshift galaxies and minihalos," *Mon. Not. Roy. Astron. Soc.* **377** (2007) 667–676, arXiv:astro-ph/0612004.
[80] E. V. Linder, "Cosmic growth history and expansion history," *Phys. Rev. D* **72** (2005) 043529, arXiv:astro-ph/0507263.
[81] D. Polarski and R. Gannouji, "On the growth of linear perturbations," *Phys. Lett. B* **660** (2008) 439–443, arXiv:0710.1510 [astro-ph].
[82] J. A. Peacock, *Cosmological Physics.* Cambridge University Press, 1999.
[83] W. H. Press and P. Schechter, "Formation of galaxies and clusters of galaxies by self-similar gravitational condensation," *Astrophys. J.* **187** (1974) 425–438.
[84] S. Cole and C. G. Lacey, "The structure of dark matter halos in hierarchical clustering models," *Mon. Not. Roy. Astron. Soc.* **281** (1996) 716, arXiv:astro-ph/9510147.
[85] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay, "The statistics of peaks of gaussian random fields," *Astrophys. J.* **304** (1986) 15–61.
[86] P. J. E. Peebles, "Primeval adiabatic perturbations: Effect of massive neutrinos," *Astrophys. J.* **258** (1982) 415–424.
[87] D. N. Schramm and G. Steigman, "Relic neutrinos and the density of the universe," *Astrophys. J.* **243** (1981) 1.
[88] J. Huchra, M. Davis, D. Latham, and J. Tonry, "A survey of galaxy redshifts: 4. The data," *Astrophys. J. Suppl.* **52** (1983) L89–L119.
[89] S. D. M. White, C. S. Frenk, and M. Davis, "Clustering in a neutrino dominated universe," *Astrophys. J. Lett.* **274** (1983) L1–L5.
[90] P. Bode, J. P. Ostriker, and N. Turok, "Halo formation in warm dark matter models," *Astrophys. J.* **556** (2001) 93–107, arXiv:astro-ph/0010389.
[91] M. Viel, J. Lesgourgues, M. G. Haehnelt, S. Matarrese, and A. Riotto, "Constraining warm dark matter candidates including sterile neutrinos and light gravitinos with WMAP and the Lyman-alpha forest," *Phys. Rev. D* **71** (2005) 063534, arXiv:astro-ph/0501562.

[92] D. J. Eisenstein and W. Hu, "Baryonic features in the matter transfer function," *Astrophys. J.* **496** (1998) 605, arXiv:astro-ph/9709112.

[93] S. Dodelson, "The real problem with MOND," *Int. J. Mod. Phys. D* **20** (2011) 2749–2753, arXiv:1112.1320 [astro-ph.CO].

[94] **SDSS** Collaboration, D. J. Eisenstein et al., "Detection of the baryon acoustic peak in the large-scale correlation function of SDSS luminous red galaxies," *Astrophys. J.* **633** (2005) 560–574, arXiv:astro-ph/0501171.

[95] **2dFGRS** Collaboration, S. Cole et al., "The 2dF galaxy redshift survey: Power-spectrum analysis of the final dataset and cosmological implications," *Mon. Not. Roy. Astron. Soc.* **362** (2005) 505–534, arXiv:astro-ph/0501174.

[96] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, "The 6dF galaxy survey: Baryon acoustic oscillations and the local Hubble constant," *Monthly Notices of the Royal Astronomical Society* **416** no. 4, (Jul, 2011) 3017–3032. http://dx.doi.org/10.1111/j.1365-2966.2011.19250.x.

[97] C. Blake, E. A. Kazin, F. Beutler, T. M. Davis, D. Parkinson, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, and et al., "The Wigglez Dark Energy Survey: Mapping the distance-redshift relation with baryon acoustic oscillations," *Monthly Notices of the Royal Astronomical Society* **418** no. 3, (Oct, 2011) 1707–1724. http://dx.doi.org/10.1111/j.1365-2966.2011.19592.x.

[98] L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, A. J. Cuesta, and et al., "The clustering of galaxies in the SDSS-III baryon oscillation spectroscopic survey: Baryon acoustic oscillations in the data release 9 spectroscopic galaxy sample," *Monthly Notices of the Royal Astronomical Society* **427** no. 4, (Dec, 2012) 3435–3467. http://dx.doi.org/10.1111/j.1365-2966.2012.22066.x.

[99] **Planck** Collaboration, N. Aghanim et al., "Planck 2018 results. I. Overview and the cosmological legacy of Planck," *Astron. Astrophys.* **641** (2020) A1, arXiv:1807.06205 [astro-ph.CO].

[100] S. Chabanier, M. Millea, and N. Palanque-Delabrouille, "Matter power spectrum: From Lyα forest to CMB scales," *Mon. Not. Roy. Astron. Soc.* **489** no. 2, (2019) 2247–2253, arXiv:1905.08103 [astro-ph.CO].

[101] **DES** Collaboration, M. A. Troxel et al., "Dark Energy Survey year 1 results: Cosmological constraints from cosmic shear," *Phys. Rev. D* **98** no. 4, (2018) 043528, arXiv:1708.01538 [astro-ph.CO].

[102] V. Ir²i£ et al., "New constraints on the free-streaming of warm dark matter from intermediate and small scale Lyman-α forest data," *Phys. Rev. D* **96** no. 2, (2017) 023522, arXiv:1702.01764 [astro-ph.CO].

[103] N. Palanque-Delabrouille, C. Yèche, N. Schöneberg, J. Lesgourgues, M. Walther, S. Chabanier, and E. Armengaud, "Hints, neutrino bounds and WDM constraints From SDSS DR14 Lyman-α and Planck full-survey data," *JCAP* **04** (2020) 038, arXiv:1911.09073 [astro-ph.CO].

[104] **DES** Collaboration, E. O. Nadler et al., "Milky way satellite census. III. constraints on dark matter properties from observations of Milky Way satellite galaxies," *Phys. Rev. Lett.* **126** (2021) 091101, arXiv:2008.00022 [astro-ph.CO].

[105] M. Vogelsberger, F. Marinacci, P. Torrey, and E. Puchwein, "Cosmological simulations of galaxy formation," *Nature Rev. Phys.* **2** no. 1, (2020) 42–66, arXiv:1909.07976 [astro-ph.GA].

[106] T. Yanagida, "Horizontal gauge symmetry and masses of neutrinos," *Conf. Proc. C* **7902131** (1979) 95–99.

[107] R. N. Mohapatra and G. Senjanovic, "Neutrino mass and spontaneous parity nonconservation," *Phys. Rev. Lett.* **44** (1980) 912.

[108] **Super-Kamiokande** Collaboration, Y. Fukuda et al., "Evidence for oscillation of atmospheric neutrinos," *Phys. Rev. Lett.* **81** (1998) 1562–1567, arXiv:hep-ex/9807003.

[109] B. T. Cleveland, T. Daily, R. Davis, Jr., J. R. Distel, K. Lande, C. K. Lee, P. S. Wildenhain, and J. Ullman, "Measurement of the solar electron neutrino flux with the Homestake chlorine detector," *Astrophys. J.* **496** (1998) 505–526.

[110] **SNO** Collaboration, Q. R. Ahmad et al., "Direct evidence for neutrino flavor transformation from neutral current interactions in the Sudbury Neutrino Observatory," *Phys. Rev. Lett.* **89** (2002) 011301, arXiv:nucl-ex/0204008.

[111] L. Wolfenstein, "Neutrino oscillations in matter," *Phys. Rev. D* **17** (1978) 2369–2374.

[112] S. P. Mikheyev and A. Y. Smirnov, "Resonance amplification of oscillations in matter and spectroscopy of solar neutrinos," *Sov. J. Nucl. Phys.* **42** (1985) 913–917.

[113] P. A. N. Machado, "Neutrino properties and interactions," arXiv:2206.13449 [hep-ph].

[114] A. de Gouvea, "TASI lectures on neutrino physics," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in D* $\geqq$ *4*. 11, 2004. arXiv:hep-ph/0411274.

[115] A. D. Sakharov, "Violation of CP invariance, C asymmetry, and baryon asymmetry of the universe," *Pisma Zh. Eksp. Teor. Fiz.* **5** (1967) 32–35.

[116] J. A. Harvey, E. W. Kolb, D. B. Reiss, and S. Wolfram, "Calculation of cosmological baryon asymmetry in grand unified gauge models," *Nucl. Phys. B* **201** (1982) 16–100.

[117] D. V. Nanopoulos and S. Weinberg, "Mechanisms for cosmological baryon production," *Phys. Rev. D* **20** (1979) 2484.

[118] S. Weinberg, "Cosmological production of baryons," *Phys. Rev. Lett.* **42** (1979) 850–853.

[119] A. Ignatiev, N. Krasnikov, V. Kuzmin, and A. Tavkhelidze, "Universal CP noninvariant superweak interaction and baryon asymmetry of the universe," *Phys. Lett. B* **76** (1978) 436–438.

[120] M. Yoshimura, "Origin of cosmological baryon asymmetry," *Phys. Lett. B* **88** (1979) 294–298.

[121] H. Georgi and S. L. Glashow, "Unity of all elementary particle forces," *Phys. Rev. Lett.* **32** (1974) 438–441.

[122] H. Georgi, "The state of the art—gauge theories," *AIP Conf. Proc.* **23** (1975) 575–582.

[123] H. Fritzsch and P. Minkowski, "Unified interactions of leptons and hadrons," *Annals Phys.* **93** (1975) 193–266.

[124] **Super-Kamiokande** Collaboration, K. Abe et al., "Search for proton decay via $p \to e^+\pi^0$ and $p \to \mu^+\pi^0$ in 0.31 megaton·years exposure of the Super-Kamiokande water Cherenkov detector," *Phys. Rev. D* **95** no. 1, (2017) 012004, arXiv:1610.03597 [hep-ex].

[125] U. Amaldi, W. de Boer, and H. Furstenau, "Comparison of grand unified theories with electroweak and strong coupling constants measured at LEP," *Phys. Lett. B* **260** (1991) 447–455.

[126] P. Langacker and M. Luo, "Implications of precision electroweak experiments for M_t, ρ_0, $\sin^2\theta_W$ and grand unification," *Phys. Rev. D* **44** (1991) 817–822.

[127] S. Dimopoulos, S. Raby, and F. Wilczek, "Supersymmetry and the scale of unification," *Phys. Rev. D* **24** (1981) 1681–1683.

[128] S. Dimopoulos and H. Georgi, "Softly broken supersymmetry and SU(5)," *Nucl. Phys. B* **193** (1981) 150–162.

[129] L. E. Ibanez and G. G. Ross, "Low-energy predictions in supersymmetric grand unified theories," *Phys. Lett. B* **105** (1981) 439–442.

[130] M. Einhorn and D. Jones, "The weak mixing angle and unification mass in supersymmetric SU(5)," *Nucl. Phys. B* **196** (1982) 475–488.

[131] W. J. Marciano and G. Senjanovic, "Predictions of supersymmetric grand unified theories," *Phys. Rev. D* **25** (1982) 3092.

[132] J. R. Ellis, M. K. Gaillard, and D. V. Nanopoulos, "Baryon number generation in grand unified theories," *Phys. Lett. B* **80** (1979) 360. [Erratum: Phys. Lett. B 82, 464 (1979)].

[133] A. D. Dolgov and D. P. Kirilova, "On particle creation by a time dependent scalar field," *Sov. J. Nucl. Phys.* **51** (1990) 172–177.

[134] J. H. Traschen and R. H. Brandenberger, "Particle production during out-of-equilibrium phase transitions," *Phys. Rev. D* **42** (1990) 2491–2504.

[135] L. Kofman, A. D. Linde, and A. A. Starobinsky, "Reheating after inflation," *Phys. Rev. Lett.* **73** (1994) 3195–3198, arXiv:hep-th/9405187.

[136] L. Kofman, A. D. Linde, and A. A. Starobinsky, "Towards the theory of reheating after inflation," *Phys. Rev. D* **56** (1997) 3258–3295, arXiv:hep-ph/9704452.

[137] E. W. Kolb, A. D. Linde, and A. Riotto, "GUT baryogenesis after preheating," *Phys. Rev. Lett.* **77** (1996) 4290–4293, arXiv:hep-ph/9606260.

[138] A. Riotto, "Theories of baryogenesis," in *ICTP Summer School in High-Energy Physics and Cosmology*, pp. 326–436. 7, 1998. arXiv:hep-ph/9807454.

[139] M. Dine, P. Huet, and R. L. Singleton, Jr., "Baryogenesis at the electroweak scale," *Nucl. Phys. B* **375** (1992) 625–648.

[140] D. Bodeker, G. D. Moore, and K. Rummukainen, "Hard thermal loops and the sphaleron rate on the lattice," *Nucl. Phys. B Proc. Suppl.* **83** (2000) 583–585, arXiv:hep-lat/9909054.

[141] A. I. Bochkarev and M. E. Shaposhnikov, "Electroweak production of baryon asymmetry and upper bounds on the Higgs and top masses," *Mod. Phys. Lett. A* **2** (1987) 417.

[142] J. A. Harvey and M. S. Turner, "Cosmological baryon and lepton number in the presence of electroweak fermion number violation," *Phys. Rev. D* **42** (1990) 3344–3349.

[143] M. Fukugita and T. Yanagida, "Baryogenesis without grand unification," *Phys. Lett. B* **174** (1986) 45–47.

[144] L. Covi, E. Roulet, and F. Vissani, "CP violating decays in leptogenesis scenarios," *Phys. Lett. B* **384** (1996) 169–174, arXiv:hep-ph/9605319.

[145] S. Davidson and A. Ibarra, "A lower bound on the right-handed neutrino mass from leptogenesis," *Phys. Lett. B* **535** (2002) 25–32, arXiv:hep-ph/0202239.

[146] R. N. Mohapatra and J. W. F. Valle, "Neutrino mass and baryon number nonconservation in superstring models," *Phys. Rev. D* **34** (1986) 1642.

[147] S. Blanchet, P. S. B. Dev, and R. N. Mohapatra, "Leptogenesis with TeV scale inverse seesaw in SO(10)," *Phys. Rev. D* **82** (2010) 115025, arXiv:1010.1471 [hep-ph].

[148] A. Pilaftsis and T. E. J. Underwood, "Resonant leptogenesis," *Nucl. Phys. B* **692** (2004) 303–345, arXiv:hep-ph/0309342.

[149] F. R. Klinkhamer and N. Manton, "A saddle point solution in the Weinberg-Salam theory," *Phys. Rev. D* **30** (1984) 2212.

[150] V. Kuzmin, V. Rubakov, and M. Shaposhnikov, "On the anomalous electroweak baryon number nonconservation in the early universe," *Phys. Lett. B* **155** (1985) 36.

[151] P. B. Arnold and L. D. McLerran, "Sphalerons, small fluctuations and baryon number violation in electroweak theory," *Phys. Rev. D* **36** (1987) 581.

[152] P. B. Arnold and L. D. McLerran, "The sphaleron strikes back," *Phys. Rev. D* **37** (1988) 1020.

[153] S. Khlebnikov and M. Shaposhnikov, "The statistical theory of anomalous fermion number nonconservation," *Nucl. Phys. B* **308** (1988) 885–912.

[154] K. Kajantie, M. Laine, K. Rummukainen, and M. E. Shaposhnikov, "Is there a hot electroweak phase transition at $m_H \gtrsim m_W$?," *Phys. Rev. Lett.* **77** (1996) 2887–2890, arXiv:hep-ph/9605288.

[155] S. Weinberg, "Gauge and global symmetries at high temperature," *Phys. Rev. D* **9** (1974) 3357–3378.

[156] S. R. Coleman, "The fate of the false vacuum. 1. Semiclassical theory," *Phys. Rev. D* **15** (1977) 2929–2936. [Erratum: Phys. Rev. D 16, 1248 (1977)].

[157] C. G. Callan, Jr. and S. R. Coleman, "The fate of the false vacuum. 2. First quantum corrections," **16** (1977) 1762–1768.

[158] M. Quiros, "Finite temperature field theory and phase transitions," in *ICTP Summer School in High-Energy Physics and Cosmology*, pp. 187–259. 1, 1999. arXiv:hep-ph/9901312.

[159] G. D. Moore, "Measuring the broken phase sphaleron rate nonperturbatively," *Phys. Rev. D* **59** (1999) 014503, arXiv:hep-ph/9805264.

[160] M. Carena, M. Quiros, and C. E. M. Wagner, "Opening the window for electroweak baryogenesis," *Phys. Lett. B* **380** (1996) 81–91, arXiv:hep-ph/9603420.

[161] S. Liebler, S. Profumo, and T. Stefaniak, "Light stop mass limits from Higgs rate measurements in the MSSM: Is MSSM electroweak baryogenesis still alive after all?," *JHEP* **04** (2016) 143, arXiv:1512.09172 [hep-ph].

[162] J. M. Cline, "Is electroweak baryogenesis dead?," arXiv:1704.08911 [hep-ph].
[163] D. A. Kirzhnits, "Weinberg model in the hot universe," *JETP Lett.* **15** (1972) 529–531.
[164] D. A. Kirzhnits and A. D. Linde, "Macroscopic consequences of the Weinberg model," *Phys. Lett. B* **42** (1972) 471–474.
[165] T. W. B. Kibble, "Topology of cosmic domains and strings," *J. Phys. A* **9** (1976) 1387–1398.
[166] J. Preskill, "Magnetic monopoles," *Ann. Rev. Nucl. Part. Sci.* **34** (1984) 461–530.
[167] A. Vilenkin, "Cosmic strings and domain walls," *Phys. Rept.* **121** (1985) 263–315.
[168] Y. B. Zeldovich, I. Y. Kobzarev, and L. B. Okun, "Cosmological consequences of the spontaneous breakdown of discrete symmetry," *Zh. Eksp. Teor. Fiz.* **67** (1974) 3–11.
[169] W. H. Press, B. S. Ryden, and D. N. Spergel, "Dynamical evolution of domain walls in an expanding universe," *Astrophys. J.* **347** (1989) 590–604.
[170] A. Vilenkin, "Gravitational field of vacuum domain walls and strings," *Phys. Rev. D* **23** (1981) 852–857.
[171] G. B. Gelmini, M. Gleiser, and E. W. Kolb, "Cosmology of biased discrete symmetry breaking," *Phys. Rev. D* **39** (1989) 1558.
[172] S. E. Larsson, S. Sarkar, and P. L. White, "Evading the cosmological domain wall problem," *Phys. Rev. D* **55** (1997) 5129–5135, arXiv:hep-ph/9608319.
[173] H. B. Nielsen and P. Olesen, "Vortex line models for dual strings," *Nucl. Phys. B* **61** (1973) 45–61.
[174] A. Albrecht and N. Turok, "Evolution of cosmic string networks," *Phys. Rev. D* **40** (1989) 973–1001.
[175] D. P. Bennett and F. R. Bouchet, "High resolution simulations of cosmic string evolution. 1. Network evolution," *Phys. Rev. D* **41** (1990) 2408.
[176] B. Allen and E. P. S. Shellard, "Cosmic string evolution: A numerical simulation," *Phys. Rev. Lett.* **64** (1990) 119–122.
[177] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer, "The number of cosmic string loops," *Phys. Rev. D* **89** no. 2, (2014) 023512, arXiv:1309.6637 [astro-ph.CO].
[178] L. Lorenz, C. Ringeval, and M. Sakellariadou, "Cosmic string loop distribution on all length scales and at any redshift," *JCAP* **10** (2010) 003, arXiv:1006.0931 [astro-ph.CO].
[179] G. Vincent, N. D. Antunes, and M. Hindmarsh, "Numerical simulations of string networks in the abelian Higgs model," *Phys. Rev. Lett.* **80** (1998) 2277–2280, arXiv:hep-ph/9708427.
[180] J. J. Blanco-Pillado, K. D. Olum, and B. Shlaer, "Large parallel cosmic string simulations: New results on loop production," *Phys. Rev. D* **83** (2011) 083514, arXiv:1101.5173 [astro-ph.CO].
[181] **Planck** Collaboration, P. A. R. Ade et al., "Planck 2013 results. XXV. Searches for cosmic strings and other topological defects," *Astron. Astrophys.* **571** (2014) A25, arXiv:1303.5085 [astro-ph.CO].
[182] A. Vilenkin, "Gravitational radiation from cosmic strings," *Phys. Lett. B* **107** (1981) 47–50.
[183] R. R. Caldwell and B. Allen, "Cosmological constraints on cosmic string gravitational radiation," *Phys. Rev. D* **45** (1992) 3447–3468.
[184] T. Damour and A. Vilenkin, "Gravitational wave bursts from cosmic strings," *Phys. Rev. Lett.* **85** (2000) 3761–3764, arXiv:gr-qc/0004075.
[185] F. A. Jenet, G. B. Hobbs, W. van Straten, R. N. Manchester, M. Bailes, J. P. W. Verbiest, R. T. Edwards, A. W. Hotan, J. M. Sarkissian, and S. M. Ord, "Upper bounds on the low-frequency stochastic gravitational wave background from pulsar timing observations: Current limits and future prospects," *Astrophys. J.* **653** (2006) 1571–1576, arXiv:astro-ph/0609013.
[186] **LIGO Scientific, Virgo, KAGRA** Collaboration, R. Abbott et al., "Constraints on cosmic strings using data from the third advanced LIGO-Virgo observing run," *Phys. Rev. Lett.* **126** no. 24, (2021) 241102, arXiv:2101.12248 [gr-qc].
[187] A. Vilenkin, "Cosmic strings as gravitational lenses," *Astrophys. J. Lett.* **282** (1984) L51–L53.
[188] R. Brandenberger, H. Firouzjahi, and J. Karouby, "Lensing and CMB anisotropies by cosmic strings at a junction," *Phys. Rev. D* **77** (2008) 083502, arXiv:0710.1636 [hep-th].

[189] K. J. Mack, D. H. Wesley, and L. J. King, "Observing cosmic string loops with gravitational lensing surveys," *Phys. Rev. D* **76** (2007) 123515, arXiv:astro-ph/0702648.
[190] G. 't Hooft, "Magnetic monopoles in unified gauge theories," *Nucl. Phys. B* **79** (1974) 276–284.
[191] A. M. Polyakov, "Particle spectrum in quantum field theory," *JETP Lett.* **20** (1974) 194–195.
[192] C. P. Dokos and T. N. Tomaras, "Monopoles and dyons in the SU(5) model," *Phys. Rev. D* **21** (1980) 2940.
[193] J. Preskill, "Cosmological production of superheavy magnetic monopoles," *Phys. Rev. Lett.* **43** (1979) 1365.
[194] W. Buchmuller, P. Di Bari, and M. Plumacher, "Leptogenesis for pedestrians," *Annals Phys.* **315** (2005) 305–351, arXiv:hep-ph/0401240.
[195] M.-C. Chen, "TASI 2006 lectures on leptogenesis," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Exploring New Frontiers Using Colliders and Neutrinos*, pp. 123–176. 3, 2007. arXiv:hep-ph/0703087.
[196] J. M. Cline, "TASI lectures on early universe cosmology: Inflation, baryogenesis and dark matter," *PoS* **TASI2018** (2019) 001, arXiv:1807.08749 [hep-ph].
[197] A. A. Starobinsky, "A new type of isotropic cosmological models without singularity," *Phys. Lett. B* **91** (1980) 99–102.
[198] A. H. Guth, "The inflationary universe: A possible solution to the horizon and flatness problems," *Phys. Rev. D* **23** (1981) 347–356.
[199] A. H. Guth and E. J. Weinberg, "Could the universe have recovered from a slow first order phase transition?," *Nucl. Phys. B* **212** (1983) 321–364.
[200] A. D. Linde, "A new inflationary universe scenario: A possible solution of the horizon, flatness, homogeneity, isotropy and primordial monopole problems," *Phys. Lett. B* **108** (1982) 389–393.
[201] A. Albrecht and P. J. Steinhardt, "Cosmology for grand unified theories with radiatively induced symmetry breaking," *Phys. Rev. Lett.* **48** (1982) 1220–1223.
[202] W. H. Press, "Spontaneous production of the Zel'dovich spectrum of cosmological fluctuations," *Phys. Scripta* **21** (1980) 702.
[203] S. W. Hawking, "The development of irregularities in a single bubble inflationary universe," *Phys. Lett. B* **115** (1982) 295.
[204] A. A. Starobinsky, "Dynamics of phase transition in the new inflationary universe scenario and generation of perturbations," *Phys. Lett. B* **117** (1982) 175–178.
[205] A. H. Guth and S. Y. Pi, "Fluctuations in the new inflationary universe," *Phys. Rev. Lett.* **49** (1982) 1110–1113.
[206] J. M. Bardeen, P. J. Steinhardt, and M. S. Turner, "Spontaneous creation of almost scale-free density perturbations in an inflationary universe," *Phys. Rev. D* **28** (1983) 679.
[207] V. F. Mukhanov and G. V. Chibisov, "Quantum fluctuations and a nonsingular universe," *JETP Lett.* **33** (1981) 532–535.
[208] **Planck** Collaboration, Y. Akrami et al., "Planck 2018 results. X. Constraints on inflation," *Astron. Astrophys.* **641** (2020) A10, arXiv:1807.06211 [astro-ph.CO].
[209] M. Tristram et al., "Improved limits on the tensor-to-scalar ratio using BICEP and Planck," arXiv:2112.07961 [astro-ph.CO].
[210] A. D. Linde, "Chaotic inflation," *Phys. Lett. B* **129** (1983) 177–181.
[211] B. A. Bassett, S. Tsujikawa, and D. Wands, "Inflation dynamics and reheating," *Rev. Mod. Phys.* **78** (2006) 537–589, arXiv:astro-ph/0507632.
[212] A. Vilenkin, "The birth of inflationary universes," *Phys. Rev. D* **27** (1983) 2848.
[213] A. D. Linde, "Eternal chaotic inflation," *Mod. Phys. Lett. A* **1** (1986) 81.
[214] A. H. Guth, "Eternal inflation and its implications," *J. Phys. A* **40** (2007) 6811–6826, arXiv:hep-th/0702178.
[215] A. Vilenkin and L. H. Ford, "Gravitational effects upon cosmological phase transitions," *Phys. Rev. D* **26** (1982) 1231.
[216] A. D. Linde, "Scalar field fluctuations in expanding universe and the new inflationary universe scenario," *Phys. Lett. B* **116** (1982) 335–339.

[217] A. A. Starobinsky, "Stochastic de Sitter (inflationary) stage in the early universe," *Lect. Notes Phys.* **246** (1986) 107–126.

[218] D. Baumann, "Inflation," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics of the Large and the Small*, pp. 523–686. 2011. arXiv:0907.5424 [hep-th].

[219] D. Baumann, "Primordial cosmology," *PoS* **TASI2017** (2018) 009, arXiv:1807.03098 [hep-th].

[220] L. Boubekeur and D. H. Lyth, "Hilltop inflation," *JCAP* **07** (2005) 010, arXiv:hep-ph/0502047.

[221] K. Freese, J. A. Frieman, and A. V. Olinto, "Natural inflation with pseudo-Nambu-Goldstone bosons," *Phys. Rev. Lett.* **65** (1990) 3233–3236.

[222] F. C. Adams, J. R. Bond, K. Freese, J. A. Frieman, and A. V. Olinto, "Natural inflation: Particle physics models, power law spectra for large scale structure, and constraints from COBE," *Phys. Rev. D* **47** (1993) 426–455, arXiv:hep-ph/9207245.

[223] E. Cremmer, S. Ferrara, C. Kounnas, and D. V. Nanopoulos, "Naturally vanishing cosmological constant in N=1 supergravity," *Phys. Lett. B* **133** (1983) 61.

[224] E. Witten, "Is supersymmetry really broken?," *Int. J. Mod. Phys. A* **10** (1995) 1247–1248, arXiv:hep-th/9409111.

[225] **Supernova Search Team** Collaboration, A. G. Riess et al., "Observational evidence from supernovae for an accelerating universe and a cosmological constant," *Astron. J.* **116** (1998) 1009–1038, arXiv:astro-ph/9805201.

[226] **Supernova Cosmology Project** Collaboration, S. Perlmutter et al., "Measurements of Ω and Λ from 42 high redshift supernovae," *Astrophys. J.* **517** (1999) 565–586, arXiv:astro-ph/9812133.

[227] C. Wetterich, "Cosmology and the fate of dilatation symmetry," *Nucl. Phys. B* **302** (1988) 668–696, arXiv:1711.03844 [hep-th].

[228] B. Ratra and P. J. E. Peebles, "Cosmological consequences of a rolling homogeneous scalar field," *Phys. Rev. D* **37** (1988) 3406.

[229] P. J. E. Peebles and B. Ratra, "Cosmology with a time variable cosmological constant," *Astrophys. J. Lett.* **325** (1988) L17.

[230] J. A. Frieman, C. T. Hill, and R. Watkins, "Late time cosmological phase transitions. 1. Particle physics models and cosmic evolution," *Phys. Rev. D* **46** (1992) 1226–1238.

[231] R. R. Caldwell, R. Dave, and P. J. Steinhardt, "Cosmological imprint of an energy component with general equation of state," *Phys. Rev. Lett.* **80** (1998) 1582–1585, arXiv:astro-ph/9708069.

[232] G. Huey, L.-M. Wang, R. Dave, R. R. Caldwell, and P. J. Steinhardt, "Resolving the cosmological missing energy problem," *Phys. Rev. D* **59** (1999) 063005, arXiv:astro-ph/9804285.

[233] S. M. Carroll, "Quintessence and the rest of the world," *Phys. Rev. Lett.* **81** (1998) 3067–3070, arXiv:astro-ph/9806099.

[234] G. R. Dvali and M. Zaldarriaga, "Changing alpha with time: Implications for fifth force type experiments and quintessence," *Phys. Rev. Lett.* **88** (2002) 091303, arXiv:hep-ph/0108217.

[235] C. Armendariz-Picon, V. F. Mukhanov, and P. J. Steinhardt, "A dynamical solution to the problem of a small cosmological constant and late time cosmic acceleration," *Phys. Rev. Lett.* **85** (2000) 4438–4441, arXiv:astro-ph/0004134.

[236] M. Malquarti, E. J. Copeland, and A. R. Liddle, "K-essence and the coincidence problem," *Phys. Rev. D* **68** (2003) 023512, arXiv:astro-ph/0304277.

[237] S. Dodelson, M. Kaplinghat, and E. Stewart, "Solving the coincidence problem: tracking oscillating energy," *Phys. Rev. Lett.* **85** (2000) 5276–5279, arXiv:astro-ph/0002360.

[238] J. Khoury and A. Weltman, "Chameleon fields: Awaiting surprises for tests of gravity in space," *Phys. Rev. Lett.* **93** (2004) 171104, arXiv:astro-ph/0309300.

[239] P. Brax, C. van de Bruck, A.-C. Davis, J. Khoury, and A. Weltman, "Detecting dark energy in orbit: The cosmological chameleon," *Phys. Rev. D* **70** (2004) 123518, arXiv:astro-ph/0408415.

[240] **DES** Collaboration, T. M. C. Abbott et al., "First cosmology results using type Ia supernovae from the Dark Energy survey: Constraints on cosmological parameters," *Astrophys. J. Lett.* **872** no. 2, (2019) L30, arXiv:1811.02374 [astro-ph.CO].

[241] **SDSS** Collaboration, M. Betoule et al., "Improved cosmological constraints from a joint analysis of the SDSS-II and SNLS supernova samples," *Astron. Astrophys.* **568** (2014) A22, arXiv:1401.4064 [astro-ph.CO].

[242] **SNLS** Collaboration, M. Sullivan et al., "SNLS3: Constraints on dark energy combining the supernova legacy survey three year data with other probes," *Astrophys. J.* **737** (2011) 102, arXiv:1104.1444 [astro-ph.CO].

[243] **Supernova Cosmology Project** Collaboration, N. Suzuki et al., "The Hubble Space Telescope cluster supernova survey: V. Improving the dark energy constraints above z>1 and building an early-type-hosted supernova sample," *Astrophys. J.* **746** (2012) 85, arXiv:1105.3470 [astro-ph.CO].

[244] **Pan-STARRS1** Collaboration, D. M. Scolnic et al., "The complete light-curve sample of spectroscopically confirmed SNe Ia from Pan-STARRS1 and cosmological constraints from the combined pantheon sample," *Astrophys. J.* **859** no. 2, (2018) 101, arXiv:1710.00845 [astro-ph.CO].

[245] F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson, "The 6dF galaxy survey: Baryon acoustic oscillations and the local Hubble constant," *Mon. Not. Roy. Astron. Soc.* **416** (2011) 3017–3032, arXiv:1106.3366 [astro-ph.CO].

[246] A. J. Ross, L. Samushia, C. Howlett, W. J. Percival, A. Burden, and M. Manera, "The clustering of the SDSS DR7 main galaxy sample—I. A 4 per cent distance measure at $z = 0.15$," *Mon. Not. Roy. Astron. Soc.* **449** no. 1, (2015) 835–847, arXiv:1409.3242 [astro-ph.CO].

[247] **BOSS** Collaboration, S. Alam et al., "The clustering of galaxies in the completed SDSS-III baryon oscillation spectroscopic survey: Cosmological analysis of the DR12 galaxy sample," *Mon. Not. Roy. Astron. Soc.* **470** no. 3, (2017) 2617–2652, arXiv:1607.03155 [astro-ph.CO].

[248] M. Ata et al., "The clustering of the SDSS-IV extended baryon oscillation spectroscopic survey DR14 quasar sample: First measurement of baryon acoustic oscillations between redshift 0.8 and 2.2," *Mon. Not. Roy. Astron. Soc.* **473** no. 4, (2018) 4773–4794, arXiv:1705.06373 [astro-ph.CO].

[249] J. D. Barrow and F. J. Tipler, *The Anthropic Cosmological Principle*. Oxford U. Pr., Oxford, 1988.

[250] S. Weinberg, "The cosmological constant problem," *Rev. Mod. Phys.* **61** (1989) 1–23.

[251] A. Vilenkin, "Predictions from quantum cosmology," *Phys. Rev. Lett.* **74** (1995) 846–849, arXiv:gr-qc/9406010.

[252] H. Martel, P. R. Shapiro, and S. Weinberg, "Likely values of the cosmological constant," *Astrophys. J.* **492** (1998) 29, arXiv:astro-ph/9701099.

[253] R. Bousso, "TASI lectures on the cosmological constant," *Gen. Rel. Grav.* **40** (2008) 607–637, arXiv:0708.4231 [hep-th].

[254] S. M. Carroll, "The cosmological constant," *Living Rev. Rel.* **4** (2001) 1, arXiv:astro-ph/0004075.

[255] E. J. Copeland, M. Sami, and S. Tsujikawa, "Dynamics of dark energy," *Int. J. Mod. Phys. D* **15** (2006) 1753–1936, arXiv:hep-th/0603057.

[256] **HAWC** Collaboration, R. Alfaro et al., "All-particle cosmic ray energy spectrum measured by the HAWC experiment from 10 to 500 TeV," *Phys. Rev. D* **96** no. 12, (2017) 122001, arXiv:1710.00890 [astro-ph.HE].

[257] Y. A. Fomin, G. B. Khristiansen, G. B. Kulikov, V. G. Pogorely, V. I. Solovjeva, V. P. Sulakov, and A. V. Trubitsyn, "Energy spectrum of cosmic rays at energies of 5 x 10^{15}–5 x 10^{17} eV," in *International Cosmic Ray Conference*, vol. 2 of *International Cosmic Ray Conference*, p. 85. Aug., 1991.

[258] **TIBET III** Collaboration, M. Amenomori et al., "The all-particle spectrum of primary cosmic rays in the wide energy range from 10^{14} eV to 10^{17} eV observed with the

Tibet-III air-shower array," *Astrophys. J.* **678** (2008) 1165–1179, arXiv:0801.1803 [hep-ex].

[259] M. Nagano, T. Hara, Y. Hatano, N. Hayashida, S. Kawaguchi, K. Kamata, T. Kifune, and Y. Mizumoto, "Energy spectrum of primary cosmic rays between $10^{14.5}$ eV and 10^{18} eV," *J. Phys. G* **10** (1984) 1295.

[260] M. A. K. Glasmacher et al., "The cosmic ray energy spectrum between 10^{14} eV and 10^{16} eV," *Astropart. Phys.* **10** (1999) 291–302.

[261] **KASCADE** Collaboration, T. Antoni et al., "KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems," *Astropart. Phys.* **24** (2005) 1–25, arXiv:astro-ph/0505413.

[262] **KASCADE Grande** Collaboration, W. D. Apel et al., "Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande," *Phys. Rev. Lett.* **107** (2011) 171104, arXiv:1107.5885 [astro-ph.HE].

[263] **HiRes** Collaboration, R. U. Abbasi et al., "First observation of the Greisen-Zatsepin-Kuzmin suppression," *Phys. Rev. Lett.* **100** (2008) 101101, arXiv:astro-ph/0703099.

[264] *The Pierre Auger Observatory: Contributions to the 36th International Cosmic Ray Conference (ICRC 2019): Madison, Wisconsin, USA, July 24–August 1, 2019*. 9, 2019. arXiv:1909.09073 [astro-ph.HE].

[265] **Telescope Array** Collaboration, D. Ivanov, "Energy spectrum measured by the telescope array," *PoS* **ICRC2019** (2020) 298.

[266] J. R. Hoerandel, "Models of the knee in the energy spectrum of cosmic rays," *Astropart. Phys.* **21** (2004) 241–265, arXiv:astro-ph/0402356.

[267] **HiRes** Collaboration, D. J. Bird et al., "The cosmic ray energy spectrum observed by the Fly's Eye," *Astrophys. J.* **424** (1994) 491–502.

[268] D. Allard, E. Parizot, E. Khan, S. Goriely, and A. V. Olinto, "UHE nuclei propagation and the interpretation of the ankle in the cosmic-ray spectrum," *Astron. Astrophys.* **443** (2005) L29–L32, arXiv:astro-ph/0505566.

[269] K. Greisen, "End to the cosmic ray spectrum?," *Phys. Rev. Lett.* **16** (1966) 748–750.

[270] G. T. Zatsepin and V. A. Kuzmin, "Upper limit of the spectrum of cosmic rays," *JETP Lett.* **4** (1966) 78–80.

[271] E. Fermi, "On the origin of the cosmic radiation," *Phys. Rev.* **75** (1949) 1169–1174.

[272] W. I. Axford, E. Leer, and G. Skadron, "The acceleration of cosmic rays by shock waves," in *International Cosmic Ray Conference*, vol. 11 of *International Cosmic Ray Conference*, p. 132. Jan., 1977.

[273] A. R. Bell, "The acceleration of cosmic rays in shock fronts—I," *Mon. Not. Roy. Astron. Soc.* **182** (Jan., 1978) 147–156.

[274] R. D. Blandford and J. P. Ostriker, "Particle acceleration by astrophysical shocks," *Astrophysics J.* **221** (1978) L29–L32.

[275] M. Takeda et al., "Energy determination in the Akeno giant air shower array experiment," *Astropart. Phys.* **19** (2003) 447–462, arXiv:astro-ph/0209422.

[276] **Pierre Auger** Collaboration, J. Bellido, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and composition implications," *PoS* **ICRC2017** (2018) 506.

[277] **Telescope Array** Collaboration, D. Ikeda and W. Hanlon, "Hybrid measurement of the energy spectrum and composition of ultra-high energy cosmic rays by the telescope array," *PoS* **ICRC2017** (2018) 515.

[278] **Pierre Auger** Collaboration, A. Aab et al., "Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory," *JCAP* **04** (2017) 038, arXiv:1612.07155 [astro-ph.HE]. [Erratum: JCAP 03, E02 (2018)].

[279] A. M. Hillas, "The origin of ultrahigh-energy cosmic rays," *Ann. Rev. Astron. Astrophys.* **22** (1984) 425–444.

[280] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva, "On astrophysical solution to ultrahigh-energy cosmic rays," *Phys. Rev. D* **74** (2006) 043005, arXiv:hep-ph/0204357.

[281] D. Hooper, S. Sarkar, and A. M. Taylor, "The intergalactic propagation of ultrahigh energy cosmic ray nuclei," *Astropart. Phys.* **27** (2007) 199–212, arXiv:astro-ph/0608085.

[282] **HiRes** Collaboration, R. U. Abbasi et al., "Measurement of the flux of ultra high energy cosmic rays by the stereo technique," *Astropart. Phys.* **32** (2009) 53–60, arXiv:0904.4500 [astro-ph.HE].

[283] **Pierre Auger** Collaboration, A. Aab et al., "Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV," *Science* **357** no. 6537, (2017) 1266–1270, arXiv:1709.07321 [astro-ph.HE].

[284] **Telescope Array** Collaboration, R. U. Abbasi et al., "Evidence of intermediate-scale energy spectrum anisotropy of cosmic rays $E \geq 10^{19.2}$ eV with the telescope array surface detector," *Astrophys. J.* **862** no. 2, (2018) 91, arXiv:1802.05003 [astro-ph.HE].

[285] D. Hooper, I. Cholis, T. Linden, and K. Fang, "HAWC observations strongly favor pulsar interpretations of the cosmic-ray positron excess," *Phys. Rev. D* **96** no. 10, (2017) 103013, arXiv:1702.08436 [astro-ph.HE].

[286] K. Mannheim and R. Schlickeiser, "Interactions of cosmic ray nuclei," *Astron. Astrophys.* **286** (1994) 983–996, arXiv:astro-ph/9402042.

[287] J. Becker Tjus, B. Eichmann, F. Halzen, A. Kheirandish, and S. M. Saba, "High-energy neutrinos from radio galaxies," *Phys. Rev. D* **89** no. 12, (2014) 123005, arXiv:1406.0506 [astro-ph.HE].

[288] R. C. Gilmore, P. Madau, J. R. Primack, R. S. Somerville, and F. Haardt, "GeV gamma-ray attenuation and the high-redshift UV background," *Mon. Not. Roy. Astron. Soc.* **399** (2009) 1694, arXiv:0905.1144 [astro-ph.CO].

[289] T. M. Kneiske, "Gamma-ray background: A review," *Chin. J. Astron. Astrophys. Suppl.* **8** (2008) 219, arXiv:0707.2915 [astro-ph].

[290] F. W. Stecker, M. A. Malkan, and S. T. Scully, "Intergalactic photon spectra from the far IR to the UV Lyman limit for $0 < Z < 6$ and the optical depth of the universe to high energy gamma-rays," *Astrophys. J.* **648** (2006) 774–783, arXiv:astro-ph/0510449.

[291] A. De Angelis, G. Galanti, and M. Roncadelli, "Transparency of the universe to gamma rays," *Mon. Not. Roy. Astron. Soc.* **432** (2013) 3245–3249, arXiv:1302.6460 [astro-ph.HE].

[292] T. M. Venters, "Contribution to the extragalactic gamma-ray background from the cascades of very-high energy gamma rays from blazars," *Astrophys. J.* **710** (2010) 1530–1540, arXiv:1001.1363 [astro-ph.HE].

[293] J. G. Learned and S. Pakvasa, "Detecting tau-neutrino oscillations at PeV energies," *Astropart. Phys.* **3** (1995) 267–274, arXiv:hep-ph/9405296.

[294] H. Athar, M. Jezabek, and O. Yasuda, "Effects of neutrino mixing on high-energy cosmic neutrino flux," *Phys. Rev. D* **62** (2000) 103007, arXiv:hep-ph/0005104.

[295] D. V. Ahluwalia, "Ambiguity in source flux of high-energy cosmic/astrophysical neutrinos: Effects of bimaximal mixing and quantum gravity induced decoherence," *Mod. Phys. Lett. A* **16** (2001) 917–926, arXiv:hep-ph/0104316.

[296] J. F. Beacom, N. F. Bell, D. Hooper, S. Pakvasa, and T. J. Weiler, "Measuring flavor ratios of high-energy astrophysical neutrinos," *Phys. Rev. D* **68** (2003) 093005, arXiv:hep-ph/0307025. [Erratum: Phys. Rev. D 72, 019901 (2005)].

[297] E. Waxman and J. N. Bahcall, "High-energy neutrinos from astrophysical sources: An upper bound," *Phys. Rev. D* **59** (1999) 023002, arXiv:hep-ph/9807282.

[298] J. N. Bahcall and E. Waxman, "High-energy astrophysical neutrinos: The upper bound is robust," *Phys. Rev. D* **64** (2001) 023002, arXiv:hep-ph/9902383.

[299] **IceCube** Collaboration, M. G. Aartsen et al., "Observation of high-energy astrophysical neutrinos in three years of Icecube data," *Phys. Rev. Lett.* **113** (2014) 101101, arXiv:1405.5303 [astro-ph.HE].

[300] **IceCube** Collaboration, M. G. Aartsen et al., "Evidence for high-energy extraterrestrial neutrinos at the IceCube detector," *Science* **342** (2013) 1242856, arXiv:1311.5238 [astro-ph.HE].

[301] **IceCube** Collaboration, M. G. Aartsen et al., "First observation of PeV-energy neutrinos with IceCube," *Phys. Rev. Lett.* **111** (2013) 021103, arXiv:1304.5356 [astro-ph.HE].

[302] A. Mücke, R. J. Protheroe, R. Engel, J. P. Rachen, and T. Stanev, "BL Lac objects in the synchrotron proton blazar model," *Astroparticle Physics* **18** no. 6, (Mar., 2003) 593–613, arXiv:astro-ph/0206164 [astro-ph].

[303] H. Netzer, "Revisiting the unified model of active galactic nuclei," *Ann. Rev. Astron. Astrophys.* **53** (2015) 365–408, arXiv:1505.00811 [astro-ph.GA].

[304] **Fermi-LAT** Collaboration, S. Abdollahi et al., "Fermi Large Area Telescope fourth source catalog," *Astrophys. J. Suppl.* **247** no. 1, (2020) 33, arXiv:1902.10045 [astro-ph.HE].

[305] P. L. Biermann and P. A. Strittmatter, "Synchrotron emission from shock waves in active galactic nuclei," *Astrophys. J.* **322** (1987) 643–649.

[306] C. M. Urry and P. Padovani, "Unified schemes for radio-loud active galactic nuclei," *Publ. Astron. Soc. Pac.* **107** (1995) 803, arXiv:astro-ph/9506063.

[307] T. Sbarrato, G. Ghisellini, G. Tagliaferri, L. Foschini, M. Nardini, F. Tavecchio, and N. Gehrels, "Blazar candidates beyond redshift 4 observed by Swift," *Mon. Not. Roy. Astron. Soc.* **446** (2015) 2483–2489, arXiv:1410.0364 [astro-ph.HE].

[308] K. Mannheim, R. J. Protheroe, and J. P. Rachen, "On the cosmic ray bound for models of extragalactic neutrino production," *Phys. Rev. D* **63** (2001) 023003, arXiv:astro-ph/9812398.

[309] F. W. Stecker, "A note on high energy neutrinos from AGN cores," *Phys. Rev. D* **72** (2005) 107301, arXiv:astro-ph/0510537.

[310] J. K. Becker, P. L. Biermann, and W. Rhode, "The diffuse neutrino flux from FR-II radio galaxies and blazars: A source property based estimate," *Astropart. Phys.* **23** (2005) 355–368, arXiv:astro-ph/0502089.

[311] **IceCube** Collaboration, M. G. Aartsen et al., "Neutrino emission from the direction of the blazar TXS 0506+056 prior to the IceCube-170922A alert," *Science* **361** no. 6398, (2018) 147–151, arXiv:1807.08794 [astro-ph.HE].

[312] **IceCube, Fermi-LAT, MAGIC, AGILE, ASAS-SN, HAWC, H.E.S.S., INTEGRAL, Kanata, Kiso, Kapteyn, Liverpool Telescope, Subaru, Swift NuSTAR, VERITAS, VLA/17B-403** Collaboration, M. G. Aartsen et al., "Multimessenger observations of a flaring blazar coincident with high-energy neutrino IceCube-170922A," *Science* **361** no. 6398, (2018) eaat1378, arXiv:1807.08816 [astro-ph.HE].

[313] D. Smith, D. Hooper, and A. Vieregg, "Revisiting AGN as the source of IceCube's diffuse neutrino flux," *JCAP* **03** (2021) 031, arXiv:2007.12706 [astro-ph.HE].

[314] D. Hooper, T. Linden, and A. Vieregg, "Active galactic nuclei and the origin of Icecube's diffuse neutrino flux," *JCAP* **02** (2019) 012, arXiv:1810.02823 [astro-ph.HE].

[315] **IceCube** Collaboration, M. G. Aartsen et al., "The contribution of Fermi-2LAC blazars to the diffuse TeV-PeV neutrino flux," *Astrophys. J.* **835** no. 1, (2017) 45, arXiv:1611.03874 [astro-ph.HE].

[316] **IceCube** Collaboration, M. G. Aartsen et al., "Searches for extended and point-like neutrino sources with four years of IceCube data," *Astrophys. J.* **796** no. 2, (2014) 109, arXiv:1406.6757 [astro-ph.HE].

[317] T. Piran, "The physics of gamma-ray bursts," *Rev. Mod. Phys.* **76** (2004) 1143–1210, arXiv:astro-ph/0405503.

[318] B. Zhang and P. Meszaros, "Gamma-ray bursts: Progress, problems & prospects," *Int. J. Mod. Phys. A* **19** (2004) 2385–2472, arXiv:astro-ph/0311321.

[319] P. Meszaros, "Gamma-ray bursts," *Rept. Prog. Phys.* **69** (2006) 2259–2322, arXiv:astro-ph/0605208.

[320] P. A. Mazzali, J.-S. Deng, N. Tominaga, K. Maeda, K. Nomoto, T. Matheson, K. S. Kawabata, K. Z. Stanek, and P. M. Garnavich, "The type Ic hypernova SN 2003dh/GRB 030329," *Astrophys. J. Lett.* **599** (2003) L95–L98, arXiv:astro-ph/0309555.

[321] J. Hjorth et al., "Constraints on short gamma-ray burst models with optical limits of GRB 050509b," *Astrophys. J. Lett.* **630** (2005) L117–L120, arXiv:astro-ph/0506123.

[322] J. S. Villasenor et al., "Discovery of the short gamma-ray burst GRB 050709," *Nature* **437** (2005) 855–858, arXiv:astro-ph/0510190.

[323] J. Gorosabel et al., "The short-duration GRB 050724 host galaxy in the context of the long-duration GRB hosts," *Astron. Astrophys.* **450** (2006) 87–92, arXiv:astro-ph/0510141.

[324] M. Vietri, "On the acceleration of ultrahigh-energy cosmic rays in gamma-ray bursts," *Astrophys. J.* **453** (1995) 883–889, arXiv:astro-ph/9506081.

[325] E. Waxman, "Gamma-ray bursts: Afterglow, high-energy cosmic rays, and neutrinos," *Astrophys. J. Suppl.* **127** (2000) 519–526.

[326] E. Waxman and J. N. Bahcall, "High-energy neutrinos from cosmological gamma-ray burst fireballs," *Phys. Rev. Lett.* **78** (1997) 2292–2295, arXiv:astro-ph/9701231.

[327] J. Alvarez-Muniz, F. Halzen, and D. W. Hooper, "High-energy neutrinos from gamma-ray bursts: Event rates in neutrino telescopes," *Phys. Rev. D* **62** (2000) 093015, arXiv:astro-ph/0006027.

[328] D. Guetta, D. Hooper, J. Alvarez-Muniz, F. Halzen, and E. Reuveni, "Neutrinos from individual gamma-ray bursts in the BATSE catalog," *Astropart. Phys.* **20** (2004) 429–455, arXiv:astro-ph/0302524.

[329] S. Hummer, P. Baerwald, and W. Winter, "Neutrino emission from gamma-ray burst fireballs, revised," *Phys. Rev. Lett.* **108** (2012) 231101, arXiv:1112.1076 [astro-ph.HE].

[330] **IceCube** Collaboration, M. G. Aartsen et al., "Search for prompt neutrino emission from gamma-ray bursts with IceCube," *Astrophys. J. Lett.* **805** no. 1, (2015) L5, arXiv:1412.6510 [astro-ph.HE].

[331] V. S. Berezinsky and G. T. Zatsepin, "Cosmic neutrinos of superhigh energy," *Yad. Fiz.* **11** (1970) 200–205.

[332] F. W. Stecker, "Diffuse fluxes of cosmic high-energy neutrinos," *Astrophys. J.* **228** (1979) 919–927.

[333] C. T. Hill and D. N. Schramm, "Ultrahigh-energy cosmic ray neutrinos," *Phys. Lett. B* **131** (1983) 247.

[334] R. Engel, D. Seckel, and T. Stanev, "Neutrinos from propagation of ultrahigh-energy protons," *Phys. Rev. D* **64** (2001) 093010, arXiv:astro-ph/0101216.

[335] Z. Fodor, S. D. Katz, A. Ringwald, and H. Tu, "Bounds on the cosmogenic neutrino flux," *JCAP* **11** (2003) 015, arXiv:hep-ph/0309171.

[336] D. Hooper, A. Taylor, and S. Sarkar, "The impact of heavy nuclei on the cosmogenic neutrino flux," *Astropart. Phys.* **23** (2005) 11–17, arXiv:astro-ph/0407618.

[337] L. A. Anchordoqui, H. Goldberg, D. Hooper, S. Sarkar, and A. M. Taylor, "Predictions for the cosmogenic neutrino flux in light of new data from the Pierre Auger Observatory," *Phys. Rev. D* **76** (2007) 123008, arXiv:0709.0734 [astro-ph].

[338] R. Aloisio, D. Boncioli, A. di Matteo, A. F. Grillo, S. Petrera, and F. Salamida, "Cosmogenic neutrinos and ultra-high energy cosmic ray models," *JCAP* **10** (2015) 006, arXiv:1505.04020 [astro-ph.HE].

[339] A. M. Taylor, M. Ahlers, and D. Hooper, "Indications of negative evolution for the sources of the highest energy cosmic rays," *Phys. Rev. D* **92** no. 6, (2015) 063011, arXiv:1505.06090 [astro-ph.HE].

[340] M. Ahlers and F. Halzen, "Minimal cosmogenic neutrinos," *Phys. Rev. D* **86** (2012) 083010, arXiv:1208.4181 [astro-ph.HE].

[341] K. Kotera, D. Allard, and A. V. Olinto, "Cosmogenic neutrinos: Parameter space and detectabilty from PeV to ZeV," *JCAP* **10** (2010) 013, arXiv:1009.1382 [astro-ph.HE].

[342] S. E. Woosley, A. Heger, and T. A. Weaver, "The evolution and explosion of massive stars," *Rev. Mod. Phys.* **74** (2002) 1015–1071.

[343] K. Scholberg, "Supernova neutrino detection," *Ann. Rev. Nucl. Part. Sci.* **62** (2012) 81–103, arXiv:1205.6003 [astro-ph.IM].

[344] **Fermi-LAT** Collaboration, M. Ackermann et al., "Detection of the characteristic pion-decay signature in supernova remnants," *Science* **339** (2013) 807, arXiv:1302.3307 [astro-ph.HE].

[345] **H.E.S.S.** Collaboration, F. Aharonian, "Primary particle acceleration above 100 TeV in the shell-type supernova remnant RX J1713.7-3946 with deep H.E.S.S. observations," *Astron. Astrophys.* **464** (2007) 235–243, arXiv:astro-ph/0611813.

[346] **AGILE** Collaboration, A. Giuliani et al., "Neutral pion emission from accelerated protons in the supernova remnant W44," *Astrophys. J. Lett.* **742** (2011) L30, arXiv:1111.4868 [astro-ph.HE].

[347] T. K. Gaisser, R. J. Protheroe, and T. Stanev, "Gamma-ray production in supernova remnants," *Astrophys. J.* **492** (1998) 219, arXiv:astro-ph/9609044.

[348] T. K. Gaisser, F. Halzen, and T. Stanev, "Particle astrophysics with high-energy neutrinos," *Phys. Rept.* **258** (1995) 173–236, arXiv:hep-ph/9410384. [Erratum: Phys. Rept. 271, 355–356 (1996)].

[349] J. G. Learned and K. Mannheim, "High-energy neutrino astrophysics," *Ann. Rev. Nucl. Part. Sci.* **50** (2000) 679–749.

[350] V. M. Kaspi and D. J. Helfand, "Constraining the birth events of neutron stars," *ASP Conf. Ser.* **271** (2002) 3, arXiv:astro-ph/0201183.

[351] I. Cholis, D. Hooper, and T. Linden, "A new determination of the spectra and luminosity function of gamma-ray millisecond pulsars," arXiv:1407.5583 [astro-ph.HE].

[352] T. Linden, K. Auchettl, J. Bramante, I. Cholis, K. Fang, D. Hooper, T. Karwal, and S. W. Li, "Using HAWC to discover invisible pulsars," *Phys. Rev. D* **96** no. 10, (2017) 103016, arXiv:1703.09704 [astro-ph.HE].

[353] J. K. Becker, "High-energy neutrinos in the context of multimessenger physics," *Phys. Rept.* **458** (2008) 173–246, arXiv:0710.1557 [astro-ph].

[354] F. Halzen and D. Hooper, "High-energy neutrino astronomy: The cosmic ray connection," *Rept. Prog. Phys.* **65** (2002) 1025–1078, arXiv:astro-ph/0204527.

[355] A. M. Hopkins and J. F. Beacom, "On the normalisation of the cosmic star formation history," *Astrophys. J.* **651** (2006) 142–154, arXiv:astro-ph/0601463.

[356] J.-L. Gervais and B. Sakita, "Field theory interpretation of supergauges in dual models," *Nucl. Phys. B* **34** (1971) 632–639.

[357] Y. A. Golfand and E. P. Likhtman, "Extension of the algebra of poincare group generators and violation of p invariance," *JETP Lett.* **13** (1971) 323–326.

[358] D. V. Volkov and V. P. Akulov, "Possible universal neutrino interaction," *JETP Lett.* **16** (1972) 438–440.

[359] J. Wess and B. Zumino, "Supergauge transformations in four-dimensions," *Nucl. Phys. B* **70** (1974) 39–50.

[360] S. R. Coleman and J. Mandula, "All possible symmetries of the S matrix," *Phys. Rev.* **159** (1967) 1251–1256.

[361] R. Haag, J. T. Lopuszanski, and M. Sohnius, "All possible generators of supersymmetries of the S matrix," *Nucl. Phys. B* **88** (1975) 257.

[362] G. R. Farrar and P. Fayet, "Phenomenology of the production, decay, and detection of new hadronic states associated with supersymmetry," *Phys. Lett. B* **76** (1978) 575–579.

[363] P. Fayet, "Supergauge invariant extension of the Higgs mechanism and a model for the electron and its neutrino," *Nucl. Phys. B* **90** (1975) 104–124.

[364] P. Fayet, "Spontaneously broken supersymmetric theories of weak, electromagnetic and strong interactions," *Phys. Lett. B* **69** (1977) 489.

[365] A. Salam and J. A. Strathdee, "Supersymmetry and fermion number conservation," *Nucl. Phys. B* **87** (1975) 85–92.

[366] H. Pagels and J. R. Primack, "Supersymmetry, cosmology and new TeV physics," *Phys. Rev. Lett.* **48** (1982) 223.

[367] P. Hut, "Limits on masses and number of neutral weakly interacting particles," *Phys. Lett. B* **69** (1977) 85.

[368] S. Weinberg, "Upper bound on gauge fermion masses," *Phys. Rev. Lett.* **50** (1983) 387.
[369] H. Goldberg, "Constraint on the photino mass from cosmology," *Phys. Rev. Lett.* **50** (1983) 1419. [Erratum: Phys. Rev. Lett. 103, 099905 (2009)].
[370] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, and M. Srednicki, "Search for supersymmetry at the anti-p p collider," *Phys. Lett. B* **127** (1983) 233–241.
[371] J. R. Ellis, J. S. Hagelin, D. V. Nanopoulos, K. A. Olive, and M. Srednicki, "Supersymmetric relics from the Big Bang," *Nucl. Phys. B* **238** (1984) 453–476.
[372] S. P. Martin, "A supersymmetry primer," *Adv. Ser. Direct. High Energy Phys.* **18** (1998) 1–98, arXiv:hep-ph/9709356.
[373] J. R. Ellis, T. Falk, K. A. Olive, and M. Srednicki, "Calculations of neutralino-stau coannihilation channels and the cosmologically relevant region of MSSM parameter space," *Astropart. Phys.* **13** (2000) 181–213, arXiv:hep-ph/9905481. [Erratum: Astropart. Phys. 15, 413–414 (2001)].
[374] M. Cirelli, N. Fornengo, and A. Strumia, "Minimal dark matter," *Nucl. Phys. B* **753** (2006) 178–194, arXiv:hep-ph/0512090.
[375] J. Bramante, P. J. Fox, G. D. Kribs, and A. Martin, "Inelastic frontier: Discovering dark matter at high recoil energy," *Phys. Rev. D* **94** no. 11, (2016) 115026, arXiv:1608.02662 [hep-ph].
[376] J. Hisano, S. Matsumoto, and M. M. Nojiri, "Explosive dark matter annihilation," *Phys. Rev. Lett.* **92** (2004) 031303, arXiv:hep-ph/0307216.
[377] J. Hisano, S. Matsumoto, M. M. Nojiri, and O. Saito, "Non-perturbative effect on dark matter annihilation and gamma ray signature from galactic center," *Phys. Rev. D* **71** (2005) 063528, arXiv:hep-ph/0412403.
[378] M. Beneke, A. Bharucha, F. Dighera, C. Hellmann, A. Hryczuk, S. Recksiegel, and P. Ruiz-Femenia, "Relic density of wino-like dark matter in the MSSM," *JHEP* **03** (2016) 119, arXiv:1601.04718 [hep-ph].
[379] J. Hisano, S. Matsumoto, M. Nagai, O. Saito, and M. Senami, "Non-perturbative effect on thermal relic abundance of dark matter," *Phys. Lett. B* **646** (2007) 34–38, arXiv:hep-ph/0610249.
[380] A. Hryczuk, R. Iengo, and P. Ullio, "Relic densities including Sommerfeld enhancements in the MSSM," *JHEP* **03** (2011) 069, arXiv:1010.2172 [hep-ph].
[381] J. Hisano, K. Ishiwata, N. Nagata, and T. Takesako, "Direct detection of electroweak-interacting dark matter," *JHEP* **07** (2011) 005, arXiv:1104.0228 [hep-ph].
[382] R. J. Hill and M. P. Solon, "WIMP-nucleon scattering with heavy WIMP effective theory," *Phys. Rev. Lett.* **112** (2014) 211602, arXiv:1309.4092 [hep-ph].
[383] C. Cheung, L. J. Hall, D. Pinner, and J. T. Ruderman, "Prospects and blind spots for neutralino dark matter," *JHEP* **05** (2013) 100, arXiv:1211.4873 [hep-ph].
[384] P. Huang and C. E. M. Wagner, "Blind Spots for neutralino dark matter in the MSSM with an intermediate m_A," *Phys. Rev. D* **90** no. 1, (2014) 015018, arXiv:1404.0392 [hep-ph].
[385] J. Schechter and J. W. F. Valle, "Neutrino masses in SU(2) $\times$ U(1) theories," *Phys. Rev. D* **22** (1980) 2227.
[386] M. Gell-Mann, P. Ramond, and R. Slansky, "Complex spinors and unified theories," *Conf. Proc. C* **790927** (1979) 315–321, arXiv:1306.4669 [hep-th].
[387] P. Minkowski, "$\mu \to e\gamma$ at a rate of one out of 10^9 muon decays?," *Phys. Lett. B* **67** (1977) 421–428.
[388] S. Dodelson and L. M. Widrow, "Sterile-neutrinos as dark matter," *Phys. Rev. Lett.* **72** (1994) 17–20, arXiv:hep-ph/9303287.
[389] D. Notzold and G. Raffelt, "Neutrino dispersion at finite temperature and density," *Nucl. Phys. B* **307** (1988) 924–936.
[390] K. Abazajian, G. M. Fuller, and M. Patel, "Sterile neutrino hot, warm, and cold dark matter," *Phys. Rev. D* **64** (2001) 023501, arXiv:astro-ph/0101524.
[391] J. S. Bullock and M. Boylan-Kolchin, "Small-scale challenges to the ΛCDM paradigm," *Ann. Rev. Astron. Astrophys.* **55** (2017) 343–387, arXiv:1707.04256 [astro-ph.CO].

[392] S. Horiuchi, P. J. Humphrey, J. Onorbe, K. N. Abazajian, M. Kaplinghat, and S. Garrison-Kimmel, "Sterile neutrino dark matter bounds from galaxies of the local group," *Phys. Rev. D* **89** no. 2, (2014) 025017, arXiv:1311.0282 [astro-ph.CO].
[393] A. Schneider, "Astrophysical constraints on resonantly produced sterile neutrino dark matter," *JCAP* **04** (2016) 059, arXiv:1601.07553 [astro-ph.CO].
[394] B. W. Lee and R. E. Shrock, "Natural suppression of symmetry violation in gauge theories: Muon-lepton and electron lepton number nonconservation," *Phys. Rev. D* **16** (1977) 1444.
[395] V. Poulin, P. D. Serpico, and J. Lesgourgues, "A fresh look at linear cosmological constraints on a decaying dark matter component," *JCAP* **08** (2016) 036, arXiv:1606.02073 [astro-ph.CO].
[396] A. Boyarsky, D. Malyshev, A. Neronov, and O. Ruchayskiy, "Constraining DM properties with SPI," *Mon. Not. Roy. Astron. Soc.* **387** (2008) 1345, arXiv:0710.4922 [astro-ph].
[397] H. Yuksel, J. F. Beacom, and C. R. Watson, "Strong upper limits on sterile neutrino warm dark matter," *Phys. Rev. Lett.* **101** (2008) 121301, arXiv:0706.4084 [astro-ph].
[398] K. Perez, K. C. Y. Ng, J. F. Beacom, C. Hersh, S. Horiuchi, and R. Krivonos, "Almost closing the νMSM sterile neutrino dark matter window with NuSTAR," *Phys. Rev. D* **95** no. 12, (2017) 123002, arXiv:1609.00667 [astro-ph.HE].
[399] B. M. Roach, K. C. Y. Ng, K. Perez, J. F. Beacom, S. Horiuchi, R. Krivonos, and D. R. Wik, "NuSTAR tests of sterile-neutrino dark matter: New galactic bulge observations and combined impact," *Phys. Rev. D* **101** no. 10, (2020) 103011, arXiv:1908.09037 [astro-ph.HE].
[400] D. Malyshev, A. Neronov, and D. Eckert, "Constraints on 3.55 keV line emission from stacked observations of dwarf spheroidal galaxies," *Phys. Rev. D* **90** (2014) 103506, arXiv:1408.3531 [astro-ph.HE].
[401] X.-D. Shi and G. M. Fuller, "A new dark matter candidate: Nonthermal sterile neutrinos," *Phys. Rev. Lett.* **82** (1999) 2832–2835, arXiv:astro-ph/9810076.
[402] H.-S. Kang and G. Steigman, "Cosmological constraints on neutrino degeneracy," *Nucl. Phys. B* **372** (1992) 494–520.
[403] S. Esposito, G. Mangano, G. Miele, and O. Pisanti, "Big Bang nucleosynthesis: An accurate determination of light element yields," *Nucl. Phys. B* **568** (2000) 421–444, arXiv:astro-ph/9906232.
[404] M. Orito, T. Kajino, G. J. Mathews, and R. N. Boyd, "Neutrino degeneracy and decoupling: New limits from primordial nucleosynthesis and the cosmic microwave background," arXiv:astro-ph/0005446.
[405] S. Hannestad, "New constraints on neutrino physics from Boomerang data," *Phys. Rev. Lett.* **85** (2000) 4203–4206, arXiv:astro-ph/0005018.
[406] J. Lesgourgues and S. Pastor, "Cosmological implications of a relic neutrino asymmetry," *Phys. Rev. D* **60** (1999) 103521, arXiv:hep-ph/9904411.
[407] J. F. Cherry and S. Horiuchi, "Closing in on resonantly produced sterile neutrino dark matter," *Phys. Rev. D* **95** no. 8, (2017) 083015, arXiv:1701.07874 [hep-ph].
[408] J. M. Pendlebury et al., "Revised experimental upper limit on the electric dipole moment of the neutron," *Phys. Rev. D* **92** no. 9, (2015) 092003, arXiv:1509.04411 [hep-ex].
[409] R. D. Peccei and H. R. Quinn, "CP conservation in the presence of instantons," *Phys. Rev. Lett.* **38** (1977) 1440–1443.
[410] G. Grilli di Cortona, E. Hardy, J. Pardo Vega, and G. Villadoro, "The QCD axion, precisely," *JHEP* **01** (2016) 034, arXiv:1511.02867 [hep-ph].
[411] S. Weinberg, "A new light boson?," *Phys. Rev. Lett.* **40** (1978) 223–226.
[412] F. Wilczek, "Problem of strong P and T invariance in the presence of instantons," *Phys. Rev. Lett.* **40** (1978) 279–282.
[413] Y. Nambu, "Quasiparticles and gauge invariance in the theory of superconductivity," *Phys. Rev.* **117** (1960) 648–663.
[414] J. Goldstone, "Field theories with superconductor solutions," *Nuovo Cim.* **19** (1961) 154–164.
[415] P. Sikivie, "Experimental tests of the invisible axion," *Phys. Rev. Lett.* **51** (1983) 1415–1417. [Erratum: Phys. Rev. Lett. 52, 695 (1984)].

[416] G. G. Raffelt, "Astrophysical axion bounds," *Lect. Notes Phys.* **741** (2008) 51–71, arXiv:hep-ph/0611350.

[417] A. Ayala, I. Domínguez, M. Giannotti, A. Mirizzi, and O. Straniero, "Revisiting the bound on axion-photon coupling from globular clusters," *Phys. Rev. Lett.* **113** no. 19, (2014) 191302, arXiv:1406.6053 [astro-ph.SR].

[418] J. H. Chang, R. Essig, and S. D. McDermott, "Supernova 1987A constraints on sub-GeV dark sectors, millicharged particles, the QCD axion, and an axion-like particle," *JHEP* **09** (2018) 051, arXiv:1803.00993 [hep-ph].

[419] J. E. Kim, "Weak interaction singlet and strong CP invariance," *Phys. Rev. Lett.* **43** (1979) 103.

[420] M. A. Shifman, A. I. Vainshtein, and V. I. Zakharov, "Can confinement ensure natural CP invariance of strong interactions?," *Nucl. Phys. B* **166** (1980) 493–506.

[421] M. Dine, W. Fischler, and M. Srednicki, "A simple solution to the strong CP problem with a harmless axion," *Phys. Lett. B* **104** (1981) 199–202.

[422] A. R. Zhitnitsky, "On possible suppression of the axion hadron interactions. (In Russian)," *Sov. J. Nucl. Phys.* **31** (1980) 260.

[423] J. E. Kim, "Constraints on very light axions from cavity experiments," *Phys. Rev. D* **58** (1998) 055006, arXiv:hep-ph/9802220.

[424] L. Di Luzio, F. Mescia, and E. Nardi, "Redefining the axion window," *Phys. Rev. Lett.* **118** no. 3, (2017) 031801, arXiv:1610.07593 [hep-ph].

[425] L. Di Luzio, F. Mescia, and E. Nardi, "Window for preferred axion models," *Phys. Rev. D* **96** no. 7, (2017) 075003, arXiv:1705.05370 [hep-ph].

[426] S. Chang and K. Choi, "Hadronic axion window and the Big Bang nucleosynthesis," *Phys. Lett. B* **316** (1993) 51–56, arXiv:hep-ph/9306216.

[427] D. Baumann, D. Green, and B. Wallisch, "New target for cosmic axion searches," *Phys. Rev. Lett.* **117** no. 17, (2016) 171301, arXiv:1604.08614 [astro-ph.CO].

[428] J. Preskill, M. B. Wise, and F. Wilczek, "Cosmology of the invisible axion," *Phys. Lett. B* **120** (1983) 127–132.

[429] L. F. Abbott and P. Sikivie, "A cosmological bound on the invisible axion," *Phys. Lett. B* **120** (1983) 133–136.

[430] M. Dine and W. Fischler, "The not so harmless axion," *Phys. Lett. B* **120** (1983) 137–141.

[431] D. J. Gross, R. D. Pisarski, and L. G. Yaffe, "QCD and instantons at finite temperature," *Rev. Mod. Phys.* **53** (1981) 43.

[432] S. Borsanyi et al., "Calculation of the axion mass based on high-temperature lattice quantum chromodynamics," *Nature* **539** no. 7627, (2016) 69–71, arXiv:1606.07494 [hep-lat].

[433] K. J. Bae, J.-H. Huh, and J. E. Kim, "Update of axion CDM energy," *JCAP* **09** (2008) 005, arXiv:0806.0497 [hep-ph].

[434] O. Wantz and E. P. S. Shellard, "Axion cosmology revisited," *Phys. Rev. D* **82** (2010) 123508, arXiv:0910.1066 [astro-ph.CO].

[435] G. Ballesteros, J. Redondo, A. Ringwald, and C. Tamarit, "Standard Model-axion-seesaw-Higgs portal inflation. Five problems of particle physics and cosmology solved in one stroke," *JCAP* **08** (2017) 001, arXiv:1610.01639 [hep-ph].

[436] M. Tegmark, A. Aguirre, M. Rees, and F. Wilczek, "Dimensionless constants, cosmology and other dark matters," *Phys. Rev. D* **73** (2006) 023505, arXiv:astro-ph/0511774.

[437] A. D. Linde, "Generation of isothermal density perturbations in the inflationary universe," *Phys. Lett. B* **158** (1985) 375–380.

[438] D. Seckel and M. S. Turner, "Isothermal density perturbations in an axion dominated inflationary universe," *Phys. Rev. D* **32** (1985) 3178.

[439] J. Hamann, S. Hannestad, G. G. Raffelt, and Y. Y. Y. Wong, "Isocurvature forecast in the anthropic axion window," *JCAP* **06** (2009) 022, arXiv:0904.0647 [hep-ph].

[440] M. P. Hertzberg, M. Tegmark, and F. Wilczek, "Axion cosmology and the energy scale of inflation," *Phys. Rev. D* **78** (2008) 083507, arXiv:0807.1726 [astro-ph].

[441] M. Beltran, J. Garcia-Bellido, and J. Lesgourgues, "Isocurvature bounds on axions revisited," *Phys. Rev. D* **75** (2007) 103507, arXiv:hep-ph/0606107.
[442] R. L. Davis, "Cosmic axions from cosmic strings," *Phys. Lett. B* **180** (1986) 225–230.
[443] A. Dabholkar and J. M. Quashnock, "Pinning down the axion," *Nucl. Phys. B* **333** (1990) 815–832.
[444] S. Chang, C. Hagmann, and P. Sikivie, "Studies of the motion and decay of axion walls bounded by strings," *Phys. Rev. D* **59** (1999) 023505, arXiv:hep-ph/9807374.
[445] C. Hagmann, S. Chang, and P. Sikivie, "Axion radiation from strings," *Phys. Rev. D* **63** (2001) 125018, arXiv:hep-ph/0012361.
[446] T. Hiramatsu, M. Kawasaki, T. Sekiguchi, M. Yamaguchi, and J. Yokoyama, "Improved estimation of radiated axions from cosmological axionic strings," *Phys. Rev. D* **83** (2011) 123531, arXiv:1012.5502 [hep-ph].
[447] T. Hiramatsu, M. Kawasaki, K. Saikawa, and T. Sekiguchi, "Production of dark matter axions from collapse of string-wall systems," *Phys. Rev. D* **85** (2012) 105020, arXiv:1202.5851 [hep-ph]. [Erratum: Phys.Rev.D 86, 089902 (2012)].
[448] M. Kawasaki, K. Saikawa, and T. Sekiguchi, "Axion dark matter from topological defects," *Phys. Rev. D* **91** no. 6, (2015) 065014, arXiv:1412.0789 [hep-ph].
[449] V. B. . Klaer and G. D. Moore, "The dark-matter axion mass," *JCAP* **11** (2017) 049, arXiv:1708.07521 [hep-ph].
[450] M. Gorghetto, E. Hardy, and G. Villadoro, "Axions from strings: The attractive solution," *JHEP* **07** (2018) 151, arXiv:1806.04677 [hep-ph].
[451] M. Buschmann, J. W. Foster, and B. R. Safdi, "Early-universe simulations of the cosmological axion," *Phys. Rev. Lett.* **124** no. 16, (2020) 161103, arXiv:1906.00967 [astro-ph.CO].
[452] M. Gorghetto, E. Hardy, and G. Villadoro, "More axions from strings," *SciPost Phys.* **10** no. 2, (2021) 050, arXiv:2007.04990 [hep-ph].
[453] R. L. Davis and E. P. S. Shellard, "Do axions need inflation?," *Nucl. Phys. B* **324** (1989) 167–186.
[454] R. A. Battye and E. P. S. Shellard, "Global string radiation," *Nucl. Phys. B* **423** (1994) 260–304, arXiv:astro-ph/9311017.
[455] R. A. Battye and E. P. S. Shellard, "Axion string constraints," *Phys. Rev. Lett.* **73** (1994) 2954–2957, arXiv:astro-ph/9403018. [Erratum: Phys. Rev. Lett. 76, 2203–2204 (1996)].
[456] D. Harari and P. Sikivie, "On the evolution of global strings in the early universe," *Phys. Lett. B* **195** (1987) 361–365.
[457] C. Hagmann and P. Sikivie, "Computer simulations of the motion and decay of global strings," *Nucl. Phys. B* **363** (1991) 247–280.
[458] M. Buschmann, J. W. Foster, A. Hook, A. Peterson, D. E. Willcox, W. Zhang, and B. R. Safdi, "Dark matter from axion strings with adaptive mesh refinement," arXiv:2108.05368 [hep-ph].
[459] S. M. Barr and J. E. Kim, "New confining force solution of the QCD axion domain-wall problem," *Phys. Rev. Lett.* **113** no. 24, (2014) 241301, arXiv:1407.4311 [hep-ph].
[460] M. Pospelov, A. Ritz, and M. B. Voloshin, "Secluded WIMP dark matter," *Phys. Lett. B* **662** (2008) 53–61, arXiv:0711.4866 [hep-ph].
[461] A. Berlin, D. Hooper, and G. Krnjaic, "Thermal dark matter from a highly decoupled sector," *Phys. Rev. D* **94** no. 9, (2016) 095019, arXiv:1609.02555 [hep-ph].
[462] A. Berlin, D. Hooper, and G. Krnjaic, "PeV-scale dark matter as a thermal relic of a decoupled sector," *Phys. Lett. B* **760** (2016) 106–111, arXiv:1602.08490 [hep-ph].
[463] J. A. Evans, S. Gori, and J. Shelton, "Looking for the WIMP next door," *JHEP* **02** (2018) 100, arXiv:1712.03974 [hep-ph].
[464] P. Adshead, Y. Cui, and J. Shelton, "Chilly dark sectors and asymmetric reheating," *JHEP* **06** (2016) 016, arXiv:1604.02458 [hep-ph].
[465] P. J. Fox, "TASI lectures on WIMPs and supersymmetry," *PoS* **TASI2018** (2019) 005.

[466] K. N. Abazajian, "Neutrinos in astrophysics and cosmology: Theoretical Advanced Study Institute (TASI) 2020 lectures," in *Theoretical Advanced Study Institute in Elementary Particle Physics: The Obscure Universe: Neutrinos and Other Dark Matters*. 2, 2021. arXiv:2102.10183 [hep-ph].

[467] A. Kusenko, "Sterile neutrinos: The dark side of the light fermions," *Phys. Rept.* **481** (2009) 1–28, arXiv:0906.2968 [hep-ph].

[468] D. J. E. Marsh, "Axions and ALPs: A very short introduction," in *13th Patras Workshop on Axions, WIMPs and WISPs*, pp. 59–74. 2018. arXiv:1712.03018 [hep-ph].

[469] D. J. E. Marsh, "Axion cosmology," *Phys. Rept.* **643** (2016) 1–79, arXiv:1510.07633 [astro-ph.CO].

[470] J. E. Kim, "A theoretical review of axion," in *3rd International Conference on Particle Physics and the Early Universe*, pp. 43–53. 2000. arXiv:astro-ph/0002193.

[471] L. D. Duffy and K. van Bibber, "Axions as dark matter particles," *New J. Phys.* **11** (2009) 105008, arXiv:0904.3346 [hep-ph].

[472] R. J. Hill and M. P. Solon, "Standard model anatomy of WIMP dark matter direct detection II: QCD analysis and hadronic matrix elements," *Phys. Rev. D* **91** (2015) 043505, arXiv:1409.8290 [hep-ph].

[473] M. W. Goodman and E. Witten, "Detectability of certain dark matter candidates," *Phys. Rev. D* **31** (1985) 3059.

[474] I. Wasserman, "Possibility of detecting heavy neutral fermions in the galaxy," *Phys. Rev. D* **33** (1986) 2071–2078.

[475] S. P. Ahlen, F. T. Avignone, R. L. Brodzinski, A. K. Drukier, G. Gelmini, and D. N. Spergel, "Limits on cold dark matter candidates from an ultralow background germanium spectrometer," *Phys. Lett. B* **195** (1987) 603–608.

[476] D. O. Caldwell, R. M. Eisberg, D. M. Grumm, M. S. Witherell, B. Sadoulet, F. S. Goulding, and A. R. Smith, "Laboratory limits on galactic cold dark matter," *Phys. Rev. Lett.* **61** (1988) 510.

[477] **LZ** Collaboration, J. Aalbers et al., "First dark matter search results from the LUX-ZEPLIN (LZ) experiment," arXiv:2207.03764 [hep-ex].

[478] **CRESST** Collaboration, A. H. Abdelhameed et al., "First results from the CRESST-III low-mass dark matter program," *Phys. Rev. D* **100** no. 10, (2019) 102002, arXiv:1904.00498 [astro-ph.CO].

[479] **SuperCDMS** Collaboration, R. Agnese et al., "Search for low-mass dark matter with CDMSlite using a profile likelihood fit," *Phys. Rev. D* **99** no. 6, (2019) 062001, arXiv:1808.09098 [astro-ph.CO].

[480] **DarkSide** Collaboration, P. Agnes et al., "DarkSide-50 532-day dark matter search with low-radioactivity argon," *Phys. Rev. D* **98** no. 10, (2018) 102006, arXiv:1802.07198 [astro-ph.CO].

[481] **XENON** Collaboration, E. Aprile et al., "Dark matter search results from a one ton-year exposure of XENON1T," *Phys. Rev. Lett.* **121** no. 11, (2018) 111302, arXiv:1805.12562 [astro-ph.CO].

[482] **XENON** Collaboration, E. Aprile et al., "Light dark matter search with ionization signals in XENON1T," *Phys. Rev. Lett.* **123** no. 25, (2019) 251801, arXiv:1907.11485 [hep-ex].

[483] **PICO** Collaboration, C. Amole et al., "Dark matter search results from the complete exposure of the PICO-60 C_3F_8 bubble chamber," *Phys. Rev. D* **100** no. 2, (2019) 022001, arXiv:1902.04031 [astro-ph.CO].

[484] **XENON** Collaboration, E. Aprile et al., "Constraining the spin-dependent WIMP-nucleon cross sections with XENON1T," *Phys. Rev. Lett.* **122** no. 14, (2019) 141301, arXiv:1902.03234 [astro-ph.CO].

[485] J. Billard, L. Strigari, and E. Figueroa-Feliciano, "Implication of neutrino backgrounds on the reach of next generation dark matter direct detection experiments," *Phys. Rev. D* **89** no. 2, (2014) 023524, arXiv:1307.5458 [hep-ph].

[486] F. Ruppin, J. Billard, E. Figueroa-Feliciano, and L. Strigari, "Complementarity of dark matter detectors in light of the neutrino background," *Phys. Rev. D* **90** no. 8, (2014) 083510, arXiv:1408.3581 [hep-ph].

[487] P. Sikivie, "Detection rates for 'invisible' axion searches," *Phys. Rev. D* **32** (1985) 2988. [Erratum: Phys. Rev. D 36, 974 (1987)].

[488] R. Bradley, J. Clarke, D. Kinion, L. J. Rosenberg, K. van Bibber, S. Matsuki, M. Muck, and P. Sikivie, "Microwave cavity searches for dark-matter axions," *Rev. Mod. Phys.* **75** (2003) 777–817.

[489] **ADMX** Collaboration, C. Bartram et al., "Search for 'invisible' axion dark matter in the 3.3–4.2 μeV mass range," arXiv:2110.06096 [hep-ex].

[490] **ADMX** Collaboration, T. Braine et al., "Extended search for the invisible axion with the axion dark matter experiment," *Phys. Rev. Lett.* **124** no. 10, (2020) 101303, arXiv:1910.08638 [hep-ex].

[491] **ADMX** Collaboration, N. Du et al., "A search for invisible axion dark matter with the axion dark matter experiment," *Phys. Rev. Lett.* **120** no. 15, (2018) 151301, arXiv:1804.05750 [hep-ex].

[492] M. Schumann, "Direct detection of WIMP dark matter: Concepts and status," *J. Phys. G* **46** no. 10, (2019) 103003, arXiv:1903.03026 [astro-ph.CO].

[493] L. Baudis, "Direct dark matter detection: The next decade," *Phys. Dark Univ.* **1** (2012) 94–108, arXiv:1211.7222 [astro-ph.IM].

[494] D. Tucker-Smith and N. Weiner, "Inelastic dark matter," *Phys. Rev. D* **64** (2001) 043502, arXiv:hep-ph/0101138.

[495] T. R. Slatyer, "Indirect dark matter signatures in the cosmic dark ages. I. Generalizing the bound on s-wave dark matter annihilation from Planck results," *Phys. Rev. D* **93** no. 2, (2016) 023527, arXiv:1506.03811 [hep-ph].

[496] D. P. Finkbeiner, S. Galli, T. Lin, and T. R. Slatyer, "Searching for dark matter in the CMB: A compact parameterization of energy injection from new physics," *Phys. Rev. D* **85** (2012) 043522, arXiv:1109.6322 [astro-ph.CO].

[497] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, "Updated CMB constraints on dark matter annihilation cross-sections," *Phys. Rev. D* **84** (2011) 027302, arXiv:1106.1528 [astro-ph.CO].

[498] S. Galli, F. Iocco, G. Bertone, and A. Melchiorri, "CMB constraints on dark matter models with large annihilation cross-section," *Phys. Rev. D* **80** (2009) 023505, arXiv:0905.0003 [astro-ph.CO].

[499] T. R. Slatyer, N. Padmanabhan, and D. P. Finkbeiner, "CMB constraints on WIMP annihilation: Energy absorption during the recombination epoch," *Phys. Rev. D* **80** (2009) 043526, arXiv:0906.1197 [astro-ph.CO].

[500] F. W. Stecker, "The cosmic gamma-ray background from the annihilation of primordial stable neutral heavy leptons," *Astrophys. J.* **223** (1978) 1032–1036.

[501] J. E. Gunn, B. W. Lee, I. Lerche, D. N. Schramm, and G. Steigman, "Some astrophysical consequences of the existence of a heavy stable neutral lepton," *Astrophys. J.* **223** (1978) 1015–1031.

[502] **CTA** Collaboration, A. Acharyya et al., "Sensitivity of the Cherenkov Telescope Array to a dark matter signal from the galactic centre," *JCAP* **01** (2021) 057, arXiv:2007.16129 [astro-ph.HE].

[503] **H.E.S.S.** Collaboration, H. Abdallah et al., "Search for dark matter annihilations towards the inner galactic halo from 10 years of observations with H.E.S.S.," *Phys. Rev. Lett.* **117** no. 11, (2016) 111301, arXiv:1607.08142 [astro-ph.HE].

[504] **VERITAS** Collaboration, S. Archambault et al., "Dark matter constraints from a joint analysis of dwarf spheroidal galaxy observations with VERITAS," *Phys. Rev. D* **95** no. 8, (2017) 082001, arXiv:1703.04937 [astro-ph.HE].

[505] **MAGIC, Fermi-LAT** Collaboration, M. L. Ahnen et al., "Limits to dark matter annihilation cross-section from a combined analysis of MAGIC and Fermi-LAT observations of dwarf satellite galaxies," *JCAP* **02** (2016) 039, arXiv:1601.06590 [astro-ph.HE].

[506] **HAWC** Collaboration, A. U. Abeysekara et al., "A search for dark matter in the galactic halo with HAWC," *JCAP* **02** (2018) 049, arXiv:1710.10288 [astro-ph.HE].

[507] C. Blanco, J. P. Harding, and D. Hooper, "Novel gamma-ray signatures of PeV-scale dark matter," *JCAP* **04** (2018) 060, arXiv:1712.02805 [hep-ph].

[508] T. Sjostrand, S. Mrenna, and P. Z. Skands, "PYTHIA 6.4 physics and manual," *JHEP* **05** (2006) 026, arXiv:hep-ph/0603175.

[509] A. V. Belikov and D. Hooper, "The contribution of inverse compton scattering to the diffuse extragalactic gamma-ray background from annihilating dark matter," *Phys. Rev. D* **81** (2010) 043505, arXiv:0906.2251 [astro-ph.CO].

[510] S. Profumo and T. E. Jeltema, "Extragalactic inverse compton light from dark matter annihilation and the Pamela positron excess," *JCAP* **07** (2009) 020, arXiv:0906.0001 [astro-ph.CO].

[511] M. Cirelli, P. D. Serpico, and G. Zaharijas, "Bremsstrahlung gamma rays from light dark matter," *JCAP* **11** (2013) 035, arXiv:1307.7152 [astro-ph.HE].

[512] J. F. Navarro, A. Ludlow, V. Springel, J. Wang, M. Vogelsberger, S. D. M. White, A. Jenkins, C. S. Frenk, and A. Helmi, "The diversity and similarity of cold dark matter halos," *Mon. Not. Roy. Astron. Soc.* **402** (2010) 21, arXiv:0810.1522 [astro-ph].

[513] J. Diemand, M. Kuhlen, P. Madau, M. Zemp, B. Moore, D. Potter, and J. Stadel, "Clumps and streams in the local dark matter distribution," *Nature* **454** (2008) 735–738, arXiv:0805.1244 [astro-ph].

[514] J. F. Navarro, C. S. Frenk, and S. D. M. White, "The Structure of cold dark matter halos," *Astrophys. J.* **462** (1996) 563–575, arXiv:astro-ph/9508025.

[515] J. F. Navarro, C. S. Frenk, and S. D. M. White, "A universal density profile from hierarchical clustering," *Astrophys. J.* **490** (1997) 493–508, arXiv:astro-ph/9611107.

[516] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. M. White, "The Aquarius project: the subhalos of galactic halos," *Mon. Not. Roy. Astron. Soc.* **391** (2008) 1685–1711, arXiv:0809.0898 [astro-ph].

[517] O. Y. Gnedin, D. Ceverino, N. Y. Gnedin, A. A. Klypin, A. V. Kravtsov, R. Levine, D. Nagai, and G. Yepes, "Halo contraction effect in hydrodynamic simulations of galaxy formation," arXiv:1108.5736 [astro-ph.CO].

[518] O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, and D. Nagai, "Response of dark matter halos to condensation of baryons: Cosmological simulations and improved adiabatic contraction model," *Astrophys. J.* **616** (2004) 16–26, arXiv:astro-ph/0406247.

[519] F. Governato, A. Zolotov, A. Pontzen, C. Christensen, S. H. Oh, A. M. Brooks, T. Quinn, S. Shen, and J. Wadsley, "Cuspy no more: How outflows affect the central dark matter and baryon distribution in lambda CDM galaxies," *Mon. Not. Roy. Astron. Soc.* **422** (2012) 1231–1240, arXiv:1202.0554 [astro-ph.CO].

[520] M. D. Weinberg and N. Katz, "Bar-driven dark halo evolution: A resolution of the cusp-core controversy," *Astrophys. J.* **580** (2002) 627–633, arXiv:astro-ph/0110632.

[521] J. A. Sellwood, "Bars and dark matter halo cores," *Astrophys. J.* **587** (2003) 638–648, arXiv:astro-ph/0210079.

[522] C. Scannapieco et al., "The Aquila comparison project: The effects of feedback and numerical methods on simulations of galaxy formation," *Mon. Not. Roy. Astron. Soc.* **423** (2012) 1726, arXiv:1112.0315 [astro-ph.GA].

[523] M. Schaller, C. S. Frenk, R. G. Bower, T. Theuns, A. Jenkins, J. Schaye, R. A. Crain, M. Furlong, C. D. Vecchia, and I. G. McCarthy, "Baryon effects on the internal structure of ΛCDM haloes in the EAGLE simulations," *Mon. Not. Roy. Astron. Soc.* **451** no. 2, (2015) 1247–1267, arXiv:1409.8617 [astro-ph.CO].

[524] A. Di Cintio, C. B. Brook, A. A. Dutton, A. V. Macciò, G. S. Stinson, and A. Knebe, "A mass-dependent density profile for dark matter haloes including the influence of galaxy formation," *Mon. Not. Roy. Astron. Soc.* **441** no. 4, (2014) 2986–2995, arXiv:1404.5959 [astro-ph.CO].

[525] A. Di Cintio, C. B. Brook, A. V. Macciò, G. S. Stinson, A. Knebe, A. A. Dutton, and J. Wadsley, "The dependence of dark matter profiles on the stellar-to-halo mass ratio: A prediction for cusps versus cores," *Mon. Not. Roy. Astron. Soc.* **437** no. 1, (2014) 415–423, arXiv:1306.0898 [astro-ph.CO].

[526] N. Bernal, L. Necib, and T. R. Slatyer, "Spherical cows in dark matter indirect detection," *JCAP* **12** (2016) 030, arXiv:1606.00433 [astro-ph.CO].

[527] **Fermi-LAT, DES** Collaboration, A. Albert et al., "Searching for dark matter annihilation in recently discovered Milky Way satellites with Fermi-LAT," *Astrophys. J.* **834** no. 2, (2017) 110, arXiv:1611.03184 [astro-ph.HE].

[528] I. Cholis, T. Linden, and D. Hooper, "A robust excess in the cosmic-ray antiproton spectrum: Implications for annihilating dark matter," *Phys. Rev. D* **99** no. 10, (2019) 103026, arXiv:1903.02549 [astro-ph.HE].

[529] F. Calore, I. Cholis, and C. Weniger, "Background model systematics for the Fermi Gev excess," *JCAP* **03** (2015) 038, arXiv:1409.0042 [astro-ph.CO].

[530] **Fermi-LAT** Collaboration, M. Ackermann et al., "The Fermi Galactic Center GeV excess and implications for dark matter," *Astrophys. J.* **840** no. 1, (2017) 43, arXiv:1704.03910 [astro-ph.HE].

[531] D. Hooper, C. Kelso, and F. S. Queiroz, "Stringent and robust constraints on the dark matter annihilation cross section from the region of the galactic center," *Astropart. Phys.* **46** (2013) 55–70, arXiv:1209.3015 [astro-ph.HE].

[532] L. Goodenough and D. Hooper, "Possible evidence for dark matter annihilation in the inner Milky Way from the Fermi Gamma Ray Space Telescope," arXiv:0910.2998 [hep-ph].

[533] D. Hooper and L. Goodenough, "Dark matter annihilation in the galactic center as seen by the Fermi Gamma Ray Space Telescope," *Phys. Lett. B* **697** (2011) 412–428, arXiv:1010.2752 [hep-ph].

[534] D. Hooper and T. Linden, "On the origin of the gamma rays from the galactic center," *Phys. Rev. D* **84** (2011) 123005, arXiv:1110.0006 [astro-ph.HE].

[535] D. Hooper and T. R. Slatyer, "Two emission mechanisms in the Fermi bubbles: A possible signal of annihilating dark matter," *Phys. Dark Univ.* **2** (2013) 118–138, arXiv:1302.6589 [astro-ph.HE].

[536] T. Daylan, D. P. Finkbeiner, D. Hooper, T. Linden, S. K. N. Portillo, N. L. Rodd, and T. R. Slatyer, "The characterization of the gamma-ray signal from the central Milky Way: A case for annihilating dark matter," *Phys. Dark Univ.* **12** (2016) 1–23, arXiv:1402.6703 [astro-ph.HE].

[537] **DES** Collaboration, K. Bechtol et al., "Eight new Milky Way companions discovered in first-year Dark Energy Survey data," *Astrophys. J.* **807** no. 1, (2015) 50, arXiv:1503.02584 [astro-ph.GA].

[538] **DES** Collaboration, A. Drlica-Wagner et al., "Eight ultra-faint galaxy candidates discovered in year two of the Dark Energy Survey," *Astrophys. J.* **813** no. 2, (2015) 109, arXiv:1508.03622 [astro-ph.GA].

[539] S. E. Koposov, V. Belokurov, G. Torrealba, and N. W. Evans, "Beasts of the southern wild: Discovery of nine ultra faint satellites in the vicinity of the Magellanic clouds," *Astrophys. J.* **805** no. 2, (2015) 130, arXiv:1503.02079 [astro-ph.GA].

[540] G. D. Martinez, "A robust determination of Milky Way satellite properties using hierarchical mass modelling," *Mon. Not. Roy. Astron. Soc.* **451** no. 3, (2015) 2524–2535, arXiv:1309.2641 [astro-ph.GA].

[541] V. Bonnivard, C. Combet, D. Maurin, and M. G. Walker, "Spherical Jeans analysis for dark matter indirect detection in dwarf spheroidal galaxies—Impact of physical parameters

and triaxiality," *Mon. Not. Roy. Astron. Soc.* **446** (2015) 3002–3021, arXiv:1407.7822 [astro-ph.HE].

[542] A. Geringer-Sameth, S. M. Koushiappas, and M. Walker, "Dwarf galaxy annihilation and decay emission profiles for dark matter experiments," *Astrophys. J.* **801** no. 2, (2015) 74, arXiv:1408.0002 [astro-ph.CO].

[543] J. Silk and M. Srednicki, "Cosmic ray anti-protons as a probe of a photino dominated universe," *Phys. Rev. Lett.* **53** (1984) 624.

[544] J. R. Ellis, R. A. Flores, K. Freese, S. Ritz, D. Seckel, and J. Silk, "Cosmic ray constraints on the annihilations of relic particles in the galactic halo," *Phys. Lett. B* **214** (1988) 403–412.

[545] F. W. Stecker, S. Rudaz, and T. F. Walsh, "Galactic anti-protons from photinos," *Phys. Rev. Lett.* **55** (1985) 2622–2625.

[546] M. S. Turner and F. Wilczek, "Positron line radiation from halo WIMP annihilations as a dark matter signature," *Phys. Rev. D* **42** (1990) 1001–1007.

[547] M. Kamionkowski and M. S. Turner, "A distinctive positron feature from heavy WIMP annihilations in the galactic halo," *Phys. Rev. D* **43** (1991) 1774–1780.

[548] L. Bergstrom, J. Edsjo, and P. Ullio, "Cosmic anti-protons as a probe for supersymmetric dark matter?," *Astrophys. J.* **526** (1999) 215–235, arXiv:astro-ph/9902012.

[549] I. V. Moskalenko and A. W. Strong, "Production and propagation of cosmic ray positrons and electrons," *Astrophys. J.* **493** (1998) 694–707, arXiv:astro-ph/9710124.

[550] I. V. Moskalenko, A. W. Strong, J. F. Ormes, and M. S. Potgieter, "Secondary anti-protons and propagation of cosmic rays in the galaxy and heliosphere," *Astrophys. J.* **565** (2002) 280–296, arXiv:astro-ph/0106567.

[551] **AMS** Collaboration, L. Accardo et al., "High statistics measurement of the positron fraction in primary cosmic rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station," *Phys. Rev. Lett.* **113** (2014) 121101.

[552] **PAMELA** Collaboration, O. Adriani et al., "Cosmic-ray positron energy spectrum measured by PAMELA," *Phys. Rev. Lett.* **111** (2013) 081102, arXiv:1308.0133 [astro-ph.HE].

[553] D. Hooper, P. Blasi, and P. D. Serpico, "Pulsars as the sources of high energy cosmic ray positrons," *JCAP* **01** (2009) 025, arXiv:0810.1527 [astro-ph].

[554] S. Profumo, "Dissecting cosmic-ray electron-positron data with Occam's razor: The role of known pulsars," *Central Eur. J. Phys.* **10** (2011) 1–31, arXiv:0812.4457 [astro-ph].

[555] A. U. Abeysekara et al., "The 2HWC HAWC Observatory gamma ray catalog," *Astrophys. J.* **843** no. 1, (2017) 40, arXiv:1702.02992 [astro-ph.HE].

[556] **HAWC** Collaboration, A. U. Abeysekara et al., "Extended gamma-ray sources around pulsars constrain the origin of the positron flux at earth," *Science* **358** no. 6365, (2017) 911–914, arXiv:1711.06223 [astro-ph.HE].

[557] T. Sudoh, T. Linden, and D. Hooper, "The highest energy HAWC sources are likely leptonic and powered by pulsars," arXiv:2101.11026 [astro-ph.HE].

[558] L. Bergstrom, T. Bringmann, I. Cholis, D. Hooper, and C. Weniger, "New limits on dark matter annihilation from AMS cosmic ray positron data," *Phys. Rev. Lett.* **111** (2013) 171101, arXiv:1306.3983 [astro-ph.HE].

[559] A. Ibarra, A. S. Lamperstorfer, and J. Silk, "Dark matter annihilations and decays after the AMS-02 positron measurements," *Phys. Rev. D* **89** no. 6, (2014) 063539, arXiv:1309.2570 [hep-ph].

[560] **AMS** Collaboration, M. Aguilar et al., "Antiproton flux, antiproton-to-proton flux ratio, and properties of elementary particle fluxes in primary cosmic rays measured with the Alpha Magnetic Spectrometer on the International Space Station," *Phys. Rev. Lett.* **117** no. 9, (2016) 091103.

[561] F. Donato, N. Fornengo, D. Maurin, and P. Salati, "Antiprotons in cosmic rays from neutralino annihilation," *Phys. Rev. D* **69** (2004) 063501, arXiv:astro-ph/0306207.

[562] A. Cuoco, M. Krämer, and M. Korsmeier, "Novel dark matter constraints from antiprotons in light of AMS-02," *Phys. Rev. Lett.* **118** no. 19, (2017) 191102, arXiv:1610.03071 [astro-ph.HE].

[563] G. Giesen, M. Boudaud, Y. Génolini, V. Poulin, M. Cirelli, P. Salati, and P. D. Serpico, "AMS-02 antiprotons, at last! Secondary astrophysical component and immediate implications for dark matter," *JCAP* **09** (2015) 023, arXiv:1504.04276 [astro-ph.HE].

[564] F. Donato, N. Fornengo, and P. Salati, "Anti-deuterons as a signature of supersymmetric dark matter," *Phys. Rev. D* **62** (2000) 043003, arXiv:hep-ph/9904481.

[565] A. Ibarra, D. Tran, and C. Weniger, "Indirect searches for decaying dark matter," *Int. J. Mod. Phys. A* **28** (2013) 1330040, arXiv:1307.6434 [hep-ph].

[566] C. Blanco and D. Hooper, "Constraints on decaying dark matter from the isotropic gamma-ray background," *JCAP* **03** (2019) 019, arXiv:1811.05988 [astro-ph.HE].

[567] T. Cohen, K. Murase, N. L. Rodd, B. R. Safdi, and Y. Soreq, "γ -ray constraints on decaying dark matter and implications for icecube," *Phys. Rev. Lett.* **119** no. 2, (2017) 021102, arXiv:1612.05638 [hep-ph].

[568] K. Murase, R. Laha, S. Ando, and M. Ahlers, "Testing the dark matter scenario for PeV neutrinos observed in IceCube," *Phys. Rev. Lett.* **115** no. 7, (2015) 071301, arXiv:1503.04663 [hep-ph].

[569] T. R. Slatyer, "Indirect detection of dark matter," in *Theoretical Advanced Study Institute in Elementary Particle Physics: Anticipating the Next Discoveries in Particle Physics*, pp. 297–353. 2018. arXiv:1710.05137 [hep-ph].

[570] D. Hooper, "TASI lectures on indirect searches for dark matter," *PoS* **TASI2018** (2019) 010, arXiv:1812.02029 [hep-ph].

[571] **ATLAS** Collaboration, G. Aad et al., "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC," *Phys. Lett. B* **716** (2012) 1–29, arXiv:1207.7214 [hep-ex].

[572] **CMS** Collaboration, S. Chatrchyan et al., "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC," *Phys. Lett. B* **716** (2012) 30–61, arXiv:1207.7235 [hep-ex].

[573] M. J. Strassler, "Theoretical particle physics at hadron colliders: An introduction," in *Theoretical Advanced Study Institute in Elementary Particle Physics: String Theory and its Applications: From meV to the Planck Scale*, pp. 153–242. 2010.

[574] M. D. Schwartz, *TASI Lectures on Collider Physics*, pp. 65–100. 2018. arXiv:1709.04533 [hep-ph].

Solutions to Selected Problems

Chapter 2

Problem 2.2

The total flux observed from all galaxies in an infinitely large, static universe is given by

$$F = \int \frac{dL}{dV}\frac{1}{4\pi r^2} dV = \int \frac{dL}{dV}\frac{1}{4\pi r^2}\, 4\pi r^2 dr = \frac{dL}{dV}\, t_{\text{age}}\, c.$$

Notice that each shell of volume provides the same contribution to the observed flux, leading to $F \propto t_{\text{age}}$. This is an illustration of what is known as Olber's paradox.

Problem 2.4

The quantity $A^\mu + B_\mu$ is meaningless because one cannot add a vector and a one-form. $A_{\mu\nu}B^{\mu\nu}C_{\mu\nu}$ is meaningless because both indices are repeated three times, which means nothing in this system of notation. Lastly, $A_{\mu\nu} = B_\alpha$ meaninglessly equates a tensor to a one-form.

Problem 2.5

$$\delta^\mu_\mu \equiv \sum_{\mu=0}^{3} \delta^\mu_\mu = \sum_{\mu=0}^{3} 1 = 4.$$

Problem 2.6

Starting with the $\lambda = 1, \mu = 2, \nu = 3$ component of the first expression, this can be written as:

$$\frac{\partial F_{12}}{\partial x^3} + \frac{\partial F_{23}}{\partial x^1} + \frac{\partial F_{31}}{\partial x^2} = 0$$

$$\longrightarrow \frac{\partial B_z}{\partial z} + \frac{\partial B_x}{\partial x} + \frac{\partial B_y}{\partial y} = 0,$$

which is Gauss's law for magnetism, $\nabla \cdot \vec{B} = 0$.

Similarly, the $\lambda = 0, \mu = 1, \nu = 2$; $\lambda = 0, \mu = 2, \nu = 3$; and $\lambda = 0, \mu = 3, \nu = 1$ components of the first expression reduce to the following:

$$\begin{aligned} &\frac{\partial(-E_x)}{\partial y} + \frac{\partial B_z}{dt} + \frac{\partial(E_y)}{\partial x} = 0 \\ \longrightarrow\ &\frac{\partial(E_y)}{\partial z} + \frac{\partial B_x}{dt} + \frac{\partial(E_z)}{\partial y} = 0 \\ \longrightarrow\ &\frac{\partial(-E_z)}{\partial x} + \frac{\partial B_y}{dt} + \frac{\partial(E_x)}{\partial z} = 0, \end{aligned}$$

which are equivalent to $\partial B_i/\partial t + (\nabla \times \vec{E})_i = 0$, for each of $i = x, y, z$. Together, this is Faraday's law of induction, $\nabla \times \vec{E} = -\partial \vec{B}/\partial t$.

Before we use the second expression, we need the Maxwell field strength tensor with raised indices, $F^{\mu\nu}$. We do this using Eq. 2.14, which yields

$$F^{\mu\nu} = \begin{pmatrix} 0 & -E_x & -E_y & -E_z \\ E_x & 0 & -B_z & B_y \\ E_y & B_z & 0 & -B_x \\ E_z & -B_y & B_x & 0 \end{pmatrix}.$$

Using this, the $\mu = 0$ component of the second expression can be written as

$$\begin{aligned} &\frac{\partial F^{0\nu}}{\partial x^\nu} = 4\pi J^0 \\ \longrightarrow\ &\frac{\partial F^{00}}{\partial x^0} + \frac{\partial F^{01}}{\partial x^1} + \frac{\partial F^{02}}{\partial x^2} + \frac{\partial F^{03}}{\partial x^3} = 4\pi\rho \\ \longrightarrow\ &0 + \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 4\pi\rho, \end{aligned}$$

which is equivalent to Gauss's law, $\nabla \cdot \vec{E} = 4\pi\rho$.

Lastly, the $\mu = i = 1, 2, 3$ components of the second expression yield

$$\begin{aligned} &\frac{\partial F^{i\nu}}{x^\nu} = 4\pi J_i \\ \longrightarrow\ &\frac{\partial F^{i0}}{x^0} + \frac{\partial F^{i1}}{x^1} + \frac{\partial F^{i2}}{x^2} + \frac{\partial F^{i3}}{x^3} = 4\pi J_i \\ \longrightarrow\ &-\frac{\partial E_i}{\partial t} + (\nabla \times \vec{B})_i = 4\pi J_i, \end{aligned}$$

which combined correspond to the fourth of Maxwell's equations, $\nabla \times \vec{B} = 4\pi\vec{J} + \partial\vec{E}/\partial t$.

Problem 2.9

For a collection of particles with the same energy and mass, we can use Eqs. 2.30 and 2.31 to write:

$$w = \frac{p_{\rm tot}}{\rho_{\rm tot}} = \frac{\frac{gN}{6\pi^2}\int_m^\infty \delta(E-E')\frac{(E'^2-m^2)^{3/2}}{e^{E'/T}\pm 1}dE'}{\frac{gN}{2\pi^2}\int_m^\infty \delta(E-E')\frac{(E'^2-m^2)^{1/2}}{e^{E'/T}\pm 1}E'^2 dE'} = \frac{\frac{1}{3}\frac{(E^2-m^2)^{3/2}}{e^{E/T}\pm 1}}{\frac{(E^2-m^2)^{1/2}}{e^{E/T}\pm 1}E^2} = \frac{E^2-m^2}{3E^2},$$

where N is the number of particles. Note that for radiation ($E \gg m$), we obtain $w \approx 1/3$, while for matter ($E \approx m$), $w \approx 0$.

Problem 2.11

To incorporate the cosmological constant into the Friedmann equations, we must perform the following substitution: $(8\pi/M_{\rm Pl}^2)\, T_{\mu\nu} \to (8\pi/M_{\rm Pl}^2)\, T_{\mu\nu} + \Lambda g_{\mu\nu}$. For the 00 element, this can be written as

$$\frac{8\pi}{M_{\rm Pl}^2}\, T_{00} = \frac{8\pi}{M_{\rm Pl}^2}\, g_{00} T_0^0 = \frac{8\pi}{M_{\rm Pl}^2}\, \rho \to \frac{8\pi}{M_{\rm Pl}^2}\, \rho + \Lambda.$$

Thus, in the first Friedmann equation, we substitute $\rho \to \rho + \Lambda M_{\rm Pl}^2/8\pi$, leading to

$$\left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = \frac{8\pi}{3M_{\rm Pl}^2}\left(\rho + \frac{\Lambda M_{\rm Pl}^2}{8\pi}\right),$$

which is equivalent to Eq. 2.39.

Similarly, for the 11 element, we substitute $8\pi\, T_{11}/M_{\rm Pl}^2 \to 8\pi\, T_{11}/M_{\rm Pl}^2 + \Lambda g_{11}$:

$$\frac{8\pi}{M_{\rm Pl}^2}\, T_{11} = \frac{8\pi}{M_{\rm Pl}^2}\, g_{11} T_1^1 = \frac{8\pi}{M_{\rm Pl}^2}\left(\frac{-a^2}{1-\kappa r^2}\right)\Big(-p\Big) \to \frac{8\pi}{M_{\rm Pl}^2}\left(\frac{-a^2}{1-\kappa r^2}\right)\Big(-p\Big) + \Lambda\left(\frac{-a^2}{1-\kappa r^2}\right),$$

which simplifies to $p \to p - \Lambda M_{\rm Pl}^2/8\pi$. Performing this operation on Eq. 2.28, we arrive at

$$2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a}\right)^2 + \frac{\kappa}{a^2} = -\frac{8\pi}{M_{\rm Pl}^2}\left(p - \frac{\Lambda M_{\rm Pl}^2}{8\pi}\right),$$

which, when combined with the first Friedmann equation, is equivalent to Eq. 2.40.

Problem 2.13

For a flat matter-dominated universe, the first Friedmann equation is given by

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3M_{\rm Pl}^2}\, \rho_0 a^{-3},$$

where ρ_0 is the current density. Rearranging this, we can write

$$\int_0^1 da\, a^{1/2} = \left(\frac{8\pi\rho_0}{3M_{\rm Pl}^2}\right)^{1/2} \int_0^{t_{\rm age}} dt,$$

and thus

$$t_{\rm age} = \frac{2}{3}\left(\frac{3M_{\rm Pl}^2}{8\pi\rho_0}\right)^{1/2} = \frac{2}{3H_0}.$$

Chapter 3

Problem 3.3

These quarks and antiquarks can combine to form the mesons known as pions,

$$\pi^+ = u\bar{d}, \quad \pi^- = d\bar{u}, \quad \pi^0 = u\bar{u}, d\bar{d},$$

and kaons,

$$K^0 = d\bar{s}, \quad \bar{K}^0 = s\bar{d}, \quad K^+ = u\bar{s}, \quad K^- = s\bar{u}.$$

The electric charge of each of these mesons is indicated by its superscript. There also exist neutral combinations of $u\bar{u}$, $d\bar{d}$, and $s\bar{s}$ known as η and η'.

These quarks can also combine in groups of three to form baryons with a charge of $+1$,

$$p = uud, \quad \Sigma^+ = uus, \quad \Delta = uud, \quad \Sigma^\star = uus,$$

baryons with a charge of -1,

$$\Sigma^- = dds, \quad \Xi^- = dss, \quad \Delta = ddd, \quad \Sigma^\star = dds, \quad \Xi^\star = dss, \quad \Omega^- = sss,$$

neutral baryons,

$$n = udd, \quad \Lambda = uds, \quad \Sigma^0 = uds, \quad \Xi^0 = uss, \quad \Delta = udd, \quad \Sigma^\star = uds, \quad \Xi^\star = uss,$$

and the doubly charged baryon, $\Delta = uuu$. Trios of antiquarks can be combined in the same combinations, corresponding to the equivalent antibaryons.

Notice that some combinations of quarks can be combined to form multiple types of baryons. For example, uud can form either a proton or a Δ baryon. The distinction in this case is that the proton is the spin-1/2 state of these quarks, whereas the Δ is an excited state with spin-3/2. Also notice that the symbols Δ, $\Sigma^\star$, and $\Xi^\star$ are each used to denote multiple combinations of quarks, each in a spin-3/2 configuration.

Problem 3.4

The most important diagrams for the process $qq \to qq$ are the t- and u-channel exchange of a gluon. For each vertex in either of these diagrams, there is a factor of $g_s/2$. Note that since these diagrams have a total of four incoming-plus-outgoing particles, the amplitude should be unitless. Furthermore, since the gluon is massless, the propagator does not introduce any mass scale into the problem, and thus no powers of energy or momentum are required. From this, it follows that we can roughly estimate that $|\mathcal{M}|^2 \sim (g_s/2)^4$. At 100 GeV, $g_s \sim 1.2$ and thus $|\mathcal{M}|^2 \sim 0.13$. At 1 GeV, in contrast, $g_s \sim 2.5$ and $|\mathcal{M}|^2 \sim 2.4$. Repeating this for photon exchange, we replace each factor of $g_s/2$ with $g_2 \sin\theta_W$, leading to $|\mathcal{M}|^2 \sim (g_2 \sin\theta_W)^4 \sim 0.01$.

Problem 3.5

For electron-electron scattering, the approach and the result are each similar to those of the previous problem, but with factors of $g_2 \sin\theta_W$ at each photon vertex, leading to $|\mathcal{M}|^2 \sim (g_2 \sin\theta_W)^4 \sim 0.01$. In the case of neutrino-neutrino scattering, the mass of the Z in the propagator suppresses the amplitude by $\sim (E/m_Z)^2$, leading to $|\mathcal{M}|^2 \sim (g_2/4\cos\theta)^4\,(E/m_Z)^4$, which is ~ 0.002 at $E = 100\,\text{GeV}$, but only $\sim 2 \times 10^{-11}$ at $E = 1\,\text{GeV}$.

Problem 3.10

Unlike in the case of two-body decays, each instance of a muon decay produces three particles with different values for their momenta. This calculation, including the integrals over the momenta of the final state particles, is described in Chapter 9 of Griffiths's *Introduction to Elementary Particles*.

Chapter 4

Problem 4.2

At $T = 1\,\text{MeV}$ or 10 MeV, the particles present with $m > T$ are the photons, electrons, and neutrinos, yielding $g_\star \approx 2 + (7/8) \times 4 + (7/8) \times 3 \times 2 = 10.75$. At $T = 20\,\text{GeV}$, we include all of the leptons and quarks (except for the top), along with the photon and the eight gluons, for $g_\star \approx (7/8) \times 3 \times (4+2) + (7/8) \times 5 \times 3 \times 4 + 2 + 8 \times 2 = 86.25$. At $T = 100\,\text{GeV}$, we add the W and Z, for $g_\star = 86.25 + 3 \times 2 + 3 = 95.25$, and at $T = 1\,\text{TeV}$ we include the Higgs and top quark, for $g_\star = 95.25 + 1 + (7/8) \times 3 \times 4 = 106.75$.

Problem 4.4

From entropy conservation, we can write

$$\frac{T}{T_i} = \frac{a_i}{a}\left(\frac{g_{\star,S}(T_i)}{g_{\star,S}(T)}\right)^{1/3}.$$

For example, when $a/a_i = 10^6$, we have

$$\begin{aligned} T &= 1\,\text{MeV} \times \left(\frac{106.75}{g_{\star,S}(T)}\right)^{1/3} \\ &= 1\,\text{MeV} \times \left(\frac{106.75}{g_{\star,S}(1\,\text{MeV})}\right)^{1/3}\left(\frac{g_{\star,S}(1\,\text{MeV})}{g_{\star,S}(T)}\right)^{1/3} \\ &\approx 2.16\,\text{MeV} \times \left(\frac{10.6}{g_{\star,S}(T)}\right)^{1/3} \approx 2.16\,\text{MeV}. \end{aligned}$$

Problem 4.6

From these interactions, we can deduce $\mu_{e^+} + \mu_{e^-} = \mu_{\mu^+} + \mu_{\mu^-} = 2\mu_\gamma = 3\mu_\gamma$, which in turn requires $\mu_\gamma = 0$, $\mu_{e^+} = -\mu_{e^-}$, and $\mu_{\mu^+} = -\mu_{\mu^-}$.

Problem 4.7

The annihilation rate in this case is given by $\Gamma = n_X \langle \sigma v \rangle$, which for $T \ll m_X$ is $\Gamma \approx 2(m_X T/2\pi)^{3/2} e^{-m_X/T} \langle \sigma v \rangle$. Comparing this to the Hubble rate, $H = (8\pi^3 g_\star(T) T^4/90 M_{\rm Pl}^2)^{1/2}$, we find that $\Gamma = H$ when $T \sim 4\,{\rm GeV}$.

Problem 4.8

Evaluated at $T = 200\,{\rm MeV}$, $g_{\star,S} \approx 40$, and thus the contribution to $g_{\star,S}$ from particles other than neutrinos changes from approximately $40 - (7/8) \times 6$ to 2 as the temperature drops from 200 MeV to well below m_e. Thus $T_\nu/T \approx (8/139)^{1/3} \approx 0.4$ in the $T \ll m_e$ limit.

Problem 4.10

The comoving number density at the time of decoupling is given by

$$a_{\rm dc}^3\, n_{\rm dc} = a_{\rm dc}^3\, g \left(\frac{m T_{\rm dc}}{2\pi} \right)^{3/2} e^{-m/T_{\rm dc}}.$$

After the universe has expanded to a scale factor equal to a, the comoving number density can be written as

$$\begin{aligned} a^3\, n &= a^3\, g \left(\frac{mT}{2\pi} \right)^{3/2} e^{-(m-\mu)/T} \\ &= a^3\, g \left(\frac{m}{2\pi} \frac{T_{\rm dc} a_{\rm dc}^2}{a^2} \right)^{3/2} e^{-(m - m[1-(a_{\rm dc}/a)^2])/[T_{\rm dc}(a_{\rm dc}/a)^2]} \\ &= a_{\rm dc}^3\, g \left(\frac{m T_{\rm dc}}{2\pi} \right)^{3/2} e^{-m/T_{\rm dc}}, \end{aligned}$$

where in the second line I have made the substitions $T = T_{\rm dc} \times (a_{\rm dc}/a)^2$ and $\mu = m[1 - (a_{\rm dc}/a)^2]$.

Problem 4.12

Starting with the left-hand side of Eq. 4.36,

$$\begin{aligned} \frac{dY_X}{dt} &= \frac{dn_X}{dt} \frac{1}{s} - \frac{n_X}{s^2} \frac{ds}{dt} \\ &= \frac{dn_X}{dt} \frac{1}{s} + \frac{n_X}{s^2} 3Hs, \end{aligned}$$

where in the second line I have used $s = S/a^3$ and that fact that the total entropy is conserved, $dS/dt = 0$. Equating this to the right-hand side, we have

$$\frac{dn_X}{dt} \frac{1}{s} + \frac{n_X}{s^2} 3Hs = s \langle \sigma v \rangle \left[\left(\frac{n_X^{\rm Eq}}{s} \right)^2 - \left(\frac{n_X}{s} \right)^2 \right].$$

After multiplying both sides by s, we recover Eq. 4.35.

Chapter 5

Problem 5.1

We begin by rewriting Eq. 5.33 in the system of natural units, suppressing the factors of ϵ_0, c, and $\hbar$, and relating $e = \sqrt{4\pi\alpha}$:

$$\frac{dE_e}{dt} \approx -\frac{Z(Z+1.3)e^6 n E_e}{16\pi^3\epsilon_0^3 m_e^2 c^4 \hbar}\left[\ln\left(\frac{183}{Z^{1/3}}\right)+\frac{1}{8}\right]$$

$$\approx -\frac{4Z(Z+1.3)\alpha^3 n E_e}{m_e^2}\left[\ln\left(\frac{183}{Z^{1/3}}\right)+\frac{1}{8}\right].$$

Using $\sigma_t = 8\pi\alpha^2/3m_e^2$, we find that this expression is equivalent to Eq. 5.5 so long as

$$4Z(Z+1.3)\left[\ln\left(\frac{183}{Z^{1/3}}\right)+\frac{1}{8}\right] = \left[\frac{4}{3}\phi_1^i - \frac{1}{3}\phi_2^i\right],$$

which is true to within a few percent for the values of ϕ given in the text.

Problem 5.4

In the current universe, $\rho_{\rm CMB} \approx 0.26\,{\rm eV/cm^3}$ and $T_{\rm CMB} \approx 0.00023\,{\rm eV}$. For $E_e \ll m_e^2/(2.7\,T_{\rm CMB}) \approx 420\,{\rm TeV}$ Klein-Nishina effects are unimportant and

$$\tau \equiv -E_e\left(\frac{dE_e}{dt}\right)^{-1} = \frac{3m_e^2}{4\sigma_t \rho_{\rm CMB} E_e} \approx 4.6\times10^4\,{\rm yr}.$$

For $z=10$, $\rho_{\rm CMB} \approx 0.26\,{\rm eV/cm^3}\times(1+z)^4 = 3807\,{\rm eV/cm^3}$ and $T_{\rm CMB} \approx 0.00023\,{\rm eV}\times(1+z) = 0.0025\,{\rm eV}$. Again, Klein-Nishina effects are unimportant, and $\tau = E_e\,(dE_e/dt)^{-1} \approx 3.1\,{\rm yr}$. For $z=100$, $\rho_{\rm CMB} \approx 0.26\,{\rm eV/cm^3}\times(1+z)^4 = 2.7\times10^7\,{\rm eV/cm^3}$ and $T_{\rm CMB} \approx 0.00023{\rm eV}\times(1+z) = 0.023\,{\rm eV}$. In this case, the energy loss rate is Klein-Nishina suppressed by a factor of $\sim 45m_e^4/64\pi^2 T_{\rm CMB}^2 E_e^2 \approx 9.2$, leading to

$$\tau \sim \frac{4.6\times10^4\,{\rm yr}}{(1+z)^4}\times 9.2 \approx 1.5\,{\rm days}.$$

Problem 5.5

For these two radiation targets to yield the same energy loss rate, the Klein-Nishina suppression for scattering with starlight has to be equal to a factor of $0.6/0.26 \approx 2.3$. This is the case for electrons that have the following energy:

$$E_e \sim \left(\frac{1}{2.3}\,\frac{45m_e^4}{64\pi^2 T_{\rm star}^2}\right)^{1/2} \approx 115\,{\rm TeV}.$$

Problem 5.7

Taking the proton to be moving in the x direction, the energy-momentum four-vector in this case is given by:

$$p^\mu = (E_p+\epsilon_\gamma, p_p+\epsilon_\gamma\cos\theta, \epsilon_\gamma\sin\theta, 0).$$

From this, we can find the energy in the center-of-momentum frame:

$$\begin{aligned}
E_{\rm CM}^2 &= p_\mu p^\mu = (E_p + \epsilon_\gamma)^2 - (p_p + \epsilon_\gamma \cos\theta)^2 - (\epsilon_\gamma \sin\theta)^2 \\
&= E_p^2 + \epsilon_\gamma^2 + 2E_p\epsilon_\gamma - p_p^2 - \epsilon_\gamma^2 \cos^2\theta - 2p_p\epsilon_\gamma \cos\theta - \epsilon_\gamma^2 \sin^2\theta \\
&= m_p^2 + 2E_p\epsilon_\gamma - 2p_p\epsilon_\gamma \cos\theta \\
&= m_p^2 + 2E_p\epsilon_\gamma \left(1 - \frac{p_p}{E_p}\cos\theta\right) \\
&= m_p^2 + 2E_p\epsilon_\gamma \left(1 - \frac{\gamma_p m_p v_p}{\gamma_p m_p}\cos\theta\right) \\
&= m_p^2 + 2E_p\epsilon_\gamma (1 - \beta\cos\theta).
\end{aligned}$$

Problem 5.8

In order for the invariant mass of this collision to exceed m_Δ, the following must be satisfied:

$$\begin{aligned}
E_{\rm CM}^2 &= (E_\gamma + m_p)^2 - E_\gamma^2 > m_\Delta^2 \\
\longrightarrow E_\gamma &> \frac{m_\Delta^2 - m_p^2}{2m_p} \approx 0.34\,{\rm GeV}.
\end{aligned}$$

Chapter 6

Problem 6.2

Equating the magnitudes of the gravitational and electrostatic forces, we have:

$$\begin{aligned}
\frac{GM_\odot m_e}{(1\,{\rm AU})^2} &= \frac{k_e |N_p - N_e| q_e^2}{(1\,{\rm AU})^2} \\
\longrightarrow GM_\odot m_e &= k_e A (M_\odot/m_p) q_e^2,
\end{aligned}$$

where N_p and N_e are the numbers of protons and electrons in the sun, respectively, and $A \equiv (N_p - N_e)/(N_p + N_e)$ is the charge asymmetry of the sun. Solving this equation, we obtain the following value for the charge asymmetry:

$$A = \frac{Gm_e m_p}{k_e q_e^2} \approx 4 \times 10^{-40}.$$

Problem 6.6

The main point here is that the $n_e = n_b\,(1 - Y_p)\,X_e$ expression applies only after the helium nuclei have become bound within electrically neutral atoms.

The expressions $Y_p \equiv 4n_{\rm He}/(n_{\rm H} + 4n_{\rm He})$ and $n_b = n_{\rm H} + 4n_{\rm He}$ can be combined with the principle of charge neutrality to find the total number density of electrons:

$$n_e = n_p = n_{\rm H} + 2n_{\rm He} = n_b\left(1 - \frac{Y_p}{2}\right).$$

Once the helium nuclei have all become bound within electrically neutral atoms, the number density of free electrons is given by

$$n_e = n_H\, X_e = n_b\,(1 - Y_p)\, X_e.$$

If the helium nuclei were (unrealistically) all fully ionized during the era of recombination, we would instead have

$$n_e = n_{\rm H}\, X_e + 2n_{\rm He} = n_b\,(1 - Y_p)\, X_e + \frac{n_b\, Y_p}{2},$$

which in the $X_e = 1$ limit reduces to $n_e = n_b\,(1 - Y_p/2)$.

Chapter 7

Problem 7.1

To obtain the energy of the CMB per unit energy that passes through a given area and solid angle per unit time time, we multiply Eq. 7.1 by factors of energy and $c/4\pi$,

$$\frac{dn_\gamma}{dE_\gamma} \times E \times \frac{c}{4\pi} = \frac{c}{4\pi^3}\,\frac{E_\gamma^3}{e^{E_\gamma/T} - 1},$$

which is now in units of energy per energy, per area, per time, per steradian. To obtain an expression in units of energy per frequency, per area, per time, per steradian, we multiply by a factor of h, which in natural units is simply 2π. Further dropping the factor of c, we now have an expression for the spectral radiance of the CMB:

$$B_\nu = \frac{2\pi}{4\pi^3}\,\frac{E_\gamma^3}{e^{E_\gamma/T} - 1} = \frac{2\pi}{4\pi^3}\,\frac{(2\pi\, \upsilon_\gamma)^3}{e^{2\pi\upsilon_\gamma/T} - 1} = \frac{4\pi\, \upsilon_\gamma^3}{e^{2\pi\upsilon_\gamma/T} - 1}.$$

Problem 7.5

To find the probability that a given region of the CMB will be hotter than T', we integrate over a Gaussian distribution:

$$P = \int_{T'}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(T - \langle T\rangle)^2/2\sigma^2}\, dT,$$

where $\langle T\rangle \approx 2.7255$ K and $\sigma = \langle \delta T^2\rangle^{1/2} \approx 1.1 \times 10^{-5}\, T$.

Problem 7.6

This is a trick question—sorry about that. Since the temperature variations of the CMB are so small, the fact that they are described by a Gaussian distribution is irrelevant. Instead, one simply integrates over Eq. 7.1 to determine the fraction of CMB photons that have $E_\gamma > 10^{-3}$ eV. Then combine this result with the number density of CMB photons to estimate the distance to the closest CMB photon with an energy greater than 10^{-3} eV.

Problem 7.7

Inserting the expression for $\delta T(\hat{n})/T$ from Eq. 7.7 into Eq. 7.9, we obtain the following:

$$\begin{aligned} a_{lm} &= \int d\hat{n} \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} a_{l'm'}\, Y_{l'm'}(\hat{n})\, Y^{\star}_{lm}(\hat{n}) \\ &= \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} a_{l'm'} \int d\hat{n}\; Y_{l'm'}(\hat{n})\, Y^{\star}_{lm}(\hat{n}) \\ &= \sum_{l'=0}^{\infty} \sum_{m'=-l'}^{l'} a_{l'm'}\, \delta_{ll'}\, \delta_{mm'} = a_{lm}. \end{aligned}$$

In the third line, I made use of the fact that the spherical harmonics make up an orthogonal and unit-normalized basis, $\int Y_{lm}(\hat{n})\, Y^{\star}_{l'm'}(\hat{n})\, d\hat{n} = \delta_{ll'}\, \delta_{mm'}$.

Problem 7.11

Viewed from the North Pole, a 1 km diameter object at the equator subtends the following angle:

$$\frac{\theta}{2\pi} = \frac{1\,\text{km}}{2\pi R_{\oplus}} \longrightarrow \theta = \frac{1\,\text{km}}{R_{\oplus}}.$$

In contrast, the same object viewed through Euclidean space at a distance of $2\pi R_{\oplus}/4$ would subtend an angle given by:

$$\tan\frac{\theta}{2} = \frac{0.5\,\text{km}}{2\pi R_{\oplus}/4} \longrightarrow \theta \approx \frac{1\,\text{km}}{R_{\oplus}}\frac{2}{\pi}.$$

The earth's positive curvature thus increases the angular size of this object by a factor of approximately $\pi/2$.

Problem 7.13

From Eq. 7.20 and using $\Omega_b = \Omega_M \approx 0.315$, the sound speed in the early universe is given by

$$c_s = c\sqrt{w} = \frac{c}{3^{1/2}(1+\frac{3\Omega_b a}{4\Omega_\gamma})^{1/2}} \approx \frac{c}{3^{1/2}(1+3,400\,a)^{1/2}}.$$

Using this expression for c_s and numerically integrating Eq. 7.19, we obtain a value of 0.097 Mpc for the sound horizon, which is about 30% smaller than that obtained using the standard baryon density. As a result, the locations of the acoustic peaks will be shifted to the right by a factor of approximately 1.3.

In addition, the larger baryon density would significantly enhance the magnitude of the odd-numbered peaks in the CMB's angular power spectrum (which correspond to regions that were in a state of maximum compression at the time of decoupling). Furthermore, the higher baryon density would also reduce the rate of diffusion, leading to less suppression of power on small angular scales.

Problem 7.14

The contribution of this particle to $N_{\rm eff}$ is given by the following:

$$\Delta N_{{\rm eff},X} = \frac{g_X}{2}\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm dc})}\right)^{4/3}\left(\frac{11}{4}\right)^{4/3},$$

where $g_X = 2$ and $g_{\star,S}(T) \approx 3.909$ at late times. Without any additional states in the thermal bath, $g_{\star,S}(T_{\rm dc}) \approx 106.75$, yielding $\Delta N_{{\rm eff},X} \approx 0.047$. If there were other particle species with TeV-scale masses in the thermal bath, however, this would have increased the value of $g_{\star,S}(T_{\rm dc})$, thus decreasing the contribution to $N_{\rm eff}$.

Chapter 8

Problem 8.3

Whereas the Hubble term in Eq. 8.11 is proportional to $\sim T^2$, the $n \to p$ collision term scales as $\sim T^2 \times T^{3/2}\, e^{-Q/T}$, allowing us to write

$$\frac{\Gamma_{n\to p}\, n_n}{H} = A\,\frac{T_{\rm FO}^{7/2}\, e^{-Q/T_{\rm FO}}}{T_{\rm FO}^2},$$

where A is a constant. Using the standard value of $Q = 1.293$ MeV, we find that these two terms are equal at $T_{\rm FO} \approx 0.8$ MeV for $A \approx (0.27\ {\rm MeV})^{-3/2}$.

Now, changing Q to 12.93 MeV, we can write the condition for freeze out as follows:

$$1 = \frac{\Gamma_{n\to p} n_n}{H} \approx (0.27\ {\rm MeV})^{-3/2} \times \frac{T_{\rm FO}^{7/2}\, e^{-12.93\ {\rm MeV}/T_{\rm FO}}}{T_{\rm FO}^2},$$

which yields a solution of $T_{\rm FO} \sim 5$ MeV. Thus, increasing the value of Q makes the neutron-to-proton ratio freeze out earlier. At this temperature, the neutron-to-proton ratio was $n_n/n_p \approx e^{-12.9/5} \sim 0.08$, or roughly half of the value found using the standard value of Q. This, in turn, reduces the value of the helium mass fraction to $Y_p \approx 0.15$.

Problem 8.4

Iron-56 has no spin, and thus has one spin degree of freedom. Solving the system of equations described by Eq. 8.16 including iron-56, one finds that the equilibrium mass fraction of iron is nearly unity for all $T \lesssim 0.17$ MeV.

Problem 8.5

At $T \sim 0.17$ MeV, the Hubble rate was $H \sim 0.01\ {\rm s}^{-1}$. If we take the entire baryon abundance to be in the form of helium-4, $n_{\rm He} = n_b/4 \approx \eta\, n_\gamma/4 \approx \eta\, \zeta(3) T^3/2\pi^2 \sim 2 \times 10^{19}\ {\rm cm}^{-3}$. In order to obtain equilibrium, the scattering rate of a chromium-48 nucleus with helium must exceed the Hubble rate, $\Gamma = \sigma\, v\, n_{\rm He} \gtrsim H$, which in turn requires $\sigma\, v \gtrsim 5 \times 10^{-22}\ {\rm cm}^3/{\rm s}$. This value might at first glance seem reasonable, until you realize that the cross section for this process is suppressed by coulomb repulsion by a factor of $\sim e^{-230}$ (see Eq. 8.22).

Problem 8.8

The energy density of this particle species is given by

$$\rho = \frac{7}{8}\frac{2\pi^2 (T/10)^4}{30} = \frac{7}{8}\frac{\rho_\gamma}{10^4}.$$

Using Eq. 8.26, this translates to $\Delta N_{\rm eff} = (11/4)^{4/3} \times 10^{-4} \approx 0.00039$. Because these particles are so cold, they can easily evade the constraints from BBN and the CMB.

Problem 8.9

Since $\Gamma \propto T^5$ and $H \propto \rho^{1/2}$, we write the relationship between the freeze-out temperature and $\Delta N_{\rm eff}$ as follows:

$$1 = \frac{\Gamma}{H} = A\frac{T_{\rm FO}^5}{[\Delta N_{\rm eff}(7/8)(4/11)^{4/3} + 1 + 3 \times (7/8)(4/11)^{4/3}]^{1/2}\, T_{\rm FO}^2},$$

where A is a constant. For $T_{\rm FO} \approx 0.8\,{\rm MeV}$ and $\Delta N_{\rm eff} = 0$, this yields $A = 2.5\,{\rm MeV}^{-3}$. Note that the three terms in the square brackets account for the energy density in dark radiation, photons, and neutrinos, respectively. From this, we can write the freeze-out temperature as a function of $\Delta N_{\rm eff}$:

$$\begin{aligned} T_{\rm FO} &\approx \left(\frac{[\Delta N_{\rm eff}(7/8)(4/11)^{4/3} + 1 + 3 \times (7/8)(4/11)^{4/3}]^{1/2}}{2.5}\right)^{1/3} {\rm MeV} \\ &\approx (0.8 + 0.018\,\Delta N_{\rm eff})\ {\rm MeV}, \end{aligned}$$

where in the second line, I have performed a Taylor series expansion for small $\Delta N_{\rm eff}$. Using $n_n/n_p \approx e^{-Q/T_{\rm FO}}$, we can write the helium mass fraction as follows:

$$Y_p \equiv \frac{4n_{\rm He}}{n_H + 4n_{\rm He}} = \frac{2(n_n/n_p)}{1 + (n_n/n_p)} = \frac{2e^{-Q/T_{\rm FO}}}{1 + e^{-Q/T_{\rm FO}}}.$$

For small values of $\Delta N_{\rm eff}$, this is well approximated by $Y_p \approx 0.33 + 0.01\Delta N_{\rm eff}$ (again, using a Taylor series expansion).

Chapter 9

Problem 9.1

Setting the orbiting object's centrifugal acceleration equal to its gravitational acceleration, we have

$$\frac{v^2}{r} = \frac{GM(<r)}{r^2},$$

where $M(<r)$ is the total mass enclosed within radius r. Solving for v, we have

$$v = \left(\frac{GM(<r)}{r}\right)^{1/2} = \left(\frac{G}{r}\int_0^r \rho(r)dV\right)^{1/2} \propto \left(\frac{1}{r}\int_0^r r^{-\alpha+2}dr\right)^{1/2} \propto r^{1-\alpha/2},$$

where I have parameterized the density profile as $\rho = \rho_0 r^{-\alpha}$. For the rotation curve to be flat, α must be equal to 2. To find the value of ρ_0, we integrate the density over the volume of the galaxy and set this equal to the total mass:

$$M_{\rm tot} = \int \rho(r) dV = 4\pi \int_0^\infty \rho_0 r^{-2} r^2 dr = 4\pi \rho_0 r_{\rm max},$$

and thus $\rho(r) = M_{\rm tot}/(4\pi\, r_{\rm max} r^2)$.

Problem 9.4

For a hot thermal relic that makes up 10% of the total dark matter density and that froze out when $T_f = 10\,{\rm MeV}$ ($g_{\star,S} \approx 10.7$), we can use Eq. 9.12 to determine

$$m_X \approx 18\,{\rm eV} \times \left(\frac{g_{\star,S}(T_f)}{106.75}\right) \approx 1.8\,{\rm eV}.$$

During the BBN era, these particles were highly relativistic, and contributed as follows to $N_{\rm eff}$ (see Eq. 7.30):

$$\Delta N_{{\rm eff},X} = \frac{4}{7}\left(\frac{3.90}{g_{\star,S}(T_f)}\right)^{4/3}\left(\frac{11}{4}\right)^{4/3} \approx 0.6,$$

which is in violation of the constraints derived from the light element abundances. By the time of recombination, these particles would have become non-relativistic, and thus do not behave like dark radiation for the purposes of the CMB.

Repeating this calculation for $T_f = 10\,{\rm GeV}$, the larger value of $g_{\star,S} \approx 83$ leads to $m_X \approx 14\,{\rm eV}$ and $\Delta N_{\rm eff} \approx 0.04$, consistent with all existing constraints.

Problem 9.5

For dark matter in the form of a cold thermal relic with a mass of 5 MeV, we can use Eqs. 9.17 and 9.15 to find that $x_f \approx 15.4$, and thus $T_f \approx 0.3\,{\rm MeV}$. For such a low freeze-out temperature, the energy density in the dark matter particles and their annihilation products would have significantly impacted the expansion rate during proton-neutron freeze out, increasing the primordial neutron abundance and thus leading to an unacceptably large abundance of both helium and deuterium.

Problem 9.6

Substituting $K_1(x) \approx K_2(x) \approx (\pi/2x)^{1/2}\, e^{-x}$ into Eq. 9.57, we obtain

$$\begin{aligned}\langle \sigma\, v\rangle &= \frac{1}{8 m_X^4 T}\frac{2 m_X\, e^{2m_X/T}}{\pi\; T}\int_{4m_X^2}^\infty \sigma\, s^{1/2}(s-4m_X^2)\left(\frac{\pi\; T}{2\sqrt{s}}\right)^{1/2} e^{-\sqrt{s}/T}\, ds \\ &= \frac{e^{2m_X/T}}{4\sqrt{2}\pi^{1/2} m_X^3 T^{3/2}}\int_{4m_X^2}^\infty \sigma\, s^{1/4}(s-4m_X^2)\, e^{-\sqrt{s}/T}\, ds.\end{aligned}$$

Further replacing $s \approx 4m_X^2 + v^2 m_X^2$, $\sqrt{s} \approx 2m_X + v^2 m_X/4$ (via Taylor series expansion), and $ds = 2v\, m_X^2\, dv$, this becomes

$$\begin{aligned}\langle \sigma v \rangle &= \frac{m_X}{2\sqrt{2}\pi^{1/2}T^{3/2}} \int_{4m_X^2}^{\infty} \sigma\, (4m_X^2 + v^2 m_X^2)^{1/4} v^2\, e^{-v^2 m_X/4T}\, v\, dv \\ &\approx \frac{m_X^{3/2}}{2\pi^{1/2}T^{3/2}} \int_{4m_X^2}^{\infty} \sigma\, v\, v^2\, e^{-v^2 m_X/4T}\, dv \\ &= \frac{x^{3/2}}{2\pi^{1/2}} \int_{4m_X^2}^{\infty} \sigma\, v\, v^2\, e^{-xv^2/4}\, dv.\end{aligned}$$

Problem 9.8

First, we calculate the effective number of degrees of freedom as evaluated at $x = x_f \approx 26$:

$$g_{\rm eff} \equiv g_X + g_Y(1+\Delta)^{3/2}\, e^{-x\Delta} \approx 2 + 2\,(1.1)^{3/2}\, e^{-2.6} \approx 2.17.$$

Next, we calculate the effective annihilation cross section:

$$\begin{aligned}\langle \sigma_{\rm eff} v \rangle &= \langle \sigma_{XX} v \rangle \frac{g_X^2}{g_{\rm eff}^2} + \langle \sigma_{YY} v \rangle \frac{g_Y^2 (1+\Delta)^3 e^{-2x\Delta}}{g_{\rm eff}^2} + \langle \sigma_{XY} v \rangle \frac{g_X g_Y (1+\Delta)^{3/2} e^{-x\Delta}}{g_{\rm eff}^2} \\ &\approx 10^{-27}{\rm cm}^3/{\rm s}\, \frac{2^2}{2.17^2} + 10^{-27}{\rm cm}^3/{\rm s}\, \frac{2^2 (1.1)^3 e^{-5.2}}{2.17^2} + \langle \sigma_{XY} v \rangle\, \frac{2^2 (1.1)^{3/2} e^{-2.6}}{2.17^2}.\end{aligned}$$

Setting $\langle \sigma_{\rm eff} v \rangle = 2.2 \times 10^{-26}\, {\rm cm}^3/{\rm s}$, we find that the coannihilation cross section must be $\langle \sigma_{XY} v \rangle = 2.9 \times 10^{-25}\, {\rm cm}^3/{\rm s}$ in order to obtain the desired relic abundance.

Problem 9.10

From the conservation of entropy, $a = (T_{\rm CMB}/T)(g_{\star,S}(T_{\rm CMB})/g_{\star,S}(T))^{1/3}$, it follows that

$$\begin{aligned}\frac{da}{dT} &= -\frac{T_{\rm CMB}}{T^2}\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T)}\right)^{1/3} - \frac{T_{\rm CMB}}{T}\frac{1}{3}\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T)}\right)^{-2/3} \frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}^2(T)} \frac{dg_{\star,S}(T)}{dT} \\ &= -\frac{a}{T} - \frac{a}{3g_{\star,S}(T)} \frac{dg_{\star,S}(T)}{dT}.\end{aligned}$$

Combining this with the definition of the Hubble rate, we can write

$$\begin{aligned}dt = \frac{da}{Ha} = \frac{dT}{Ha}\frac{da}{dT} &= \frac{dT}{Ha}\left[-\frac{a}{T} - \frac{a}{3g_{\star,S}(T)} \frac{dg_{\star,S}(T)}{dT} \right] \\ &= -\frac{dT}{TH}\left[1 + \frac{T}{3g_{\star,S}(T)} \frac{dg_{\star,S}(T)}{dT}\right].\end{aligned}$$

In the approximation used in the text, I dropped the second term in the square brackets. This term is most significant when the fractional change of $g_{\star,S}$ per fractional change in T

is the largest. Evaluated at most temperatures, this correction is only a few percent or less. It is most important near the QCD phase transition, where it can be as large as $\sim 50\%$.

Problem 9.13

In the large x limit, the modified Bessel functions of the second kind can be approximated by $K_1(x) \approx K_1(x) \approx (\pi/2x)^{1/2} e^{-x}$, and thus $\langle \Gamma_X \rangle \approx \Gamma_X$.

We can safely neglect the inverse decay term in this case because the number density of the decaying particles greatly exceeds their equilibrium abundance, $n_X \gg n_X^{\rm Eq}$.

Problem 9.14

From Eq. 9.60, a black hole with a temperature greater than 100 GeV (and thus a mass lighter than $\sim 10^{11}$ g) evaporates in a time τ, given by

$$\int_{M_i}^{0} M_{\rm BH}^2 \, dM_{\rm BH} = -\frac{\mathcal{G} g_{\star,H} M_{\rm Pl}^4}{30720\pi} \int_0^{\tau} dt$$

$$\rightarrow \tau = \frac{30720\pi \, M_i^3}{3\mathcal{G} g_{\star,H} M_{\rm Pl}^4} \approx 4 \times 10^{-4} \, {\rm s},$$

where $M_i = 10^8$ g is the initial mass of the black hole.

In a matter-dominated (i.e., a black hole-dominated) universe, the energy density at the evaporation time is given by $\rho = M_{\rm Pl}^2/6\pi\tau^2$, after which the temperature can be found by solving the following:

$$\rho = \frac{M_{\rm Pl}^2}{6\pi\tau^2} = \frac{\pi^2 g_\star(T) T^4}{30},$$

which yields a post-evaporation temperature of $T_{\rm PE} \approx 50$ MeV. Since this time, the scale factor of the universe has grown by a factor of $(T_{\rm PE}/T_{\rm CMB})(g_{\star,S}(T_{\rm PE})/g_{\star,S}(T_{\rm CMB}))^{1/3}$, and thus the dark matter density today is given by

$$\rho_{\rm DM} \sim 0.01 \times \frac{\pi^2 g_\star(T_{\rm PE}) \, T_{\rm PE}^4}{30} \times \left(\frac{T_{\rm CMB}^3}{T_{\rm PE}^3} \frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_{\rm PE})} \right) \sim 0.9 \, {\rm GeV/cm^3},$$

which is many orders of magnitude larger than the measured dark matter density.

If $m_X = 7 \times 10^{10}$ GeV, in contrast, the black holes do not become hot enough to produce the dark matter particles ($T \sim M_{\rm BH}$) until their masses are reduced to $M_{\rm BH} \sim 140$ g. As a result, the resulting dark matter abundance is suppressed by a factor of $\sim 140 \, {\rm g}/M_i \sim 1.4 \times 10^{-3}$, yielding $\rho_{\rm DM} \sim 1.3 \times 10^{-6} \, {\rm GeV/cm^3}$.

Chapter 10

Problem 10.3

Start by writing the temperature of the dark matter as follows:

$$T_{\rm DM} = T_{\rm kd} \, \frac{(1+z)^2}{(1+z_{\rm kd})^2}.$$

Next, by applying entropy conservation, we can express this as follows:

$$T_{\rm DM} = T_{\rm kd}\frac{(1+z)^2}{(1+z_{\rm kd})^2}\left[\frac{T}{T_{\rm kd}}\frac{(1+z_{\rm kd})}{(1+z)}\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm kd})}\right)^{1/3}\right]^{-2}\left[\frac{T_{\rm CMB}}{T_{\rm kd}}\frac{(1+z_{\rm kd})}{1}\left(\frac{g_{\star,S}(T_{\rm CMB})}{g_{\star,S}(T_{\rm kd})}\right)^{1/3}\right]^{4},$$
$$= \frac{(1+z)^4\, T_{\rm CMB}^4\, g_{\star,S}^{4/3}(T_{\rm CMB})}{T_{\rm kd}\, T^2\, g_{\star,S}^{2/3}(T)\, g_{\star,S}^{2/3}(T_{\rm kd})},$$

where each of the terms in square brackets is equal to unity. Next, writing $\bar{\rho} = \Omega_{\rm rad}\,\rho_{\rm crit}\,(1+z)^4$, we arrive at

$$\lambda_J = \left(\frac{\pi\, T_{\rm DM}}{G\bar{\rho} m_{\rm DM}}\right)^{1/2} = \left(\frac{\pi}{G\Omega_{\rm rad}\,\rho_{\rm crit}\, m_{\rm DM}}\,\frac{T_{\rm CMB}^4\, g_{\star,S}^{4/3}(T_{\rm CMB})}{T_{\rm kd}\, T^2\, g_{\star,S}^{2/3}(T)\, g_{\star,S}^{2/3}(T_{\rm kd})}\right)^{1/2}.$$

Problem 10.5

Start by taking Taylor series expansions for a and t:

$$a = \frac{4\pi G\rho_0}{3}\left[1-\left(1-\frac{\eta^2}{2}+\frac{\eta^4}{24}-...\right)\right] \approx \frac{2\pi G\rho_0}{3}\left[\eta^2-\frac{\eta^4}{12}\right]$$
$$t = \frac{4\pi G\rho_0}{3}\left[\eta-\left(\eta-\frac{\eta^3}{6}+\frac{\eta^5}{120}-...\right)\right] \approx \frac{2\pi G\rho_0}{9}\left[\eta^3-\frac{\eta^5}{20}\right].$$

From this expression for t, we can write:

$$\eta = \left[\frac{9t}{2\pi G\rho_0}\frac{1}{[1-(\eta^2/20)]}\right]^{1/3} \approx \left[\frac{9t}{2\pi G\rho_0}+\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{5/3}\right]^{1/3},$$

where in the second step I have performed another Taylor series expansion. Inserting this expression for η into the Taylor series expansion for a, we obtain the following:

$$a \approx \frac{2\pi G\rho_0}{3}\left\{\left[\frac{9t}{2\pi G\rho_0}+\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{4/3}\right]^{2/3}-\frac{1}{12}\left[\frac{9t}{2\pi G\rho_0}+\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{4/3}\right]^{4/3}\right\}$$
$$= \frac{2\pi G\rho_0}{3}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\left\{\left[1+\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right]^{2/3}\right.$$
$$\left.-\frac{1}{12}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\left[1+\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right]^{4/3}\right\}.$$

Using the binomial expansions $(1+x)^{2/3} = 1+2x/3+\cdots$ and $(1+x)^{4/3} = 1+4x/3+\cdots$, this becomes

$$a \approx \frac{2\pi G\rho_0}{3}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\left\{1+\frac{2}{3}\cdot\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right.$$

$$
\begin{aligned}
&- \frac{1}{12}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\left[1+\frac{4}{3}\cdot\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right]\Bigg\} \\
&= \frac{2\pi G\rho_0}{3}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\Bigg\{1+\left(\frac{2}{3}\cdot\frac{1}{20}-\frac{1}{12}\right)\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3} \\
&\quad - \frac{1}{12}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\frac{4}{3}\cdot\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\Bigg\} \\
&\approx \frac{2\pi G\rho_0}{3}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\left[1-\frac{1}{20}\left(\frac{9t}{2\pi G\rho_0}\right)^{2/3}\right].
\end{aligned}
$$

Problem 10.6

In this case, we can apply Eq. 10.21, but with a negative sign in front of the second term, since $H = -2/3t$ during the contracting matter-dominated phase. For a parameterization given by $\delta = At^{\alpha}$, this equation can be satisfied for $\alpha = (7 \pm \sqrt{73})/6 \approx 2.6, -0.26$, yielding a general solution of $\delta \approx A\,t^{2.6} + B\,t^{-0.26}$. The $\alpha = 2.6$ term quickly comes to dominate, leading to $\delta \propto t^{2.6}$ during the matter-dominated contracting phase.

Problem 10.8

We start with the expression for the variance,

$$
\sigma = \frac{1.686}{\sqrt{2}}\left(\frac{M}{M_\star}\right)^{-(2+n_s)/6},
$$

and take its derivative:

$$
\frac{d\sigma}{dM} = -\frac{1.686}{\sqrt{2}}\frac{(2+n_s)}{6M_\star}\left(\frac{M}{M_\star}\right)^{-(8+n_s)/6}.
$$

Inserting these expressions for σ and $d\sigma/dM$ into Eq. 10.49, we arrive at the following:

$$
\begin{aligned}
\frac{dn}{dM} &= \frac{2\bar{\rho}_M}{M}\,\frac{1.686}{\sqrt{2\pi}\sigma^2}\left|\frac{d\sigma}{dM}\right|\,\exp\left[-\frac{1.686^2}{2\sigma^2}\right] \\
&= \frac{2\bar{\rho}_M}{M}\,\frac{1}{\sqrt{\pi}}\left(\frac{M}{M_\star}\right)^{(2+n_s)/3}\frac{(2+n_s)}{6M_\star}\left(\frac{M}{M_\star}\right)^{-(8+n_s)/6}\,\exp\left[-\left(\frac{M}{M_\star}\right)^{(2+n_s)/3}\right] \\
&= \frac{2\bar{\rho}_M}{M}\,\frac{(2+n_s)}{6\sqrt{\pi}M}\left(\frac{M}{M_\star}\right)^{(2+n_s)/6}\,\exp\left[-\left(\frac{M}{M_\star}\right)^{(2+n_s)/3}\right].
\end{aligned}
$$

Problem 10.10

Using Eq. 10.59, we see that the Jeans mass depends on the temperature of kinetic decoupling only through the value of $g_{\star,S}$ evaluated at that temperature, $M_J \propto [g_{\star,S}(T_{\rm kd})]^{-1/2}$. Assuming Standard Model particle content, the value of the Jeans mass can be impacted by $T_{\rm kd}$ by no more than a factor of $[g_{\star,S}(T \gtrsim \text{TeV})/g_{\star,S}(T \lesssim 0.1\,\text{MeV})]^{1/2} \approx (106.75/3.9)^{1/2} \approx 5.2$.

Problem 10.11

Using the average energy of a relativistic particle as a proxy for temperature, $\langle E_X \rangle \sim 3T_{\rm DM}$, we can express the temperature of the dark matter as follows:

$$T_{\rm DM} \sim \frac{\langle E_X \rangle}{3}\left(\frac{1+z_{\rm nr}}{1+z_{\rm nt}}\right)\left(\frac{1+z}{1+z_{\rm nr}}\right)^2 = \frac{\langle E_X \rangle}{3}\left(\frac{T_{\rm nr}\, g_{\star,S}^{1/3}(T_{\rm nr})}{T_{\rm nt}\, g_{\star,S}^{1/3}(T_{\rm nt})}\right)\left(\frac{T\, g_{\star,S}^{1/3}(T)}{T_{\rm nr}\, g_{\star,S}^{1/3}(T_{\rm nr})}\right)^2$$
$$= \frac{\langle E_X \rangle^2}{9}\frac{T^2}{m_X T_{\rm nt}^2}\left(\frac{g_{\star,S}(T)}{g_{\star,S}(T_{\rm nt})}\right)^{2/3},$$

where the subscripts 'nt' and 'nr' denote quantities at the time or redshift of nonthermal production, and at the time that the dark matter becomes non-relativistic. In the last step, I have used the following to determine the temperature of the Standard Model bath at $z_{\rm nr}$:

$$\frac{m_X}{\langle E_X \rangle/3} \sim \frac{1+z_{\rm nr}}{1+z_{\rm nt}} = \frac{T_{\rm nr}\, g_{\star,S}^{1/3}(T_{\rm nr})}{T_{\rm nt}\, g_{\star,S}^{1/3}(T_{\rm nt})} \longrightarrow T_{\rm nr} = \frac{3 m_X T_{\rm nt}\, g_{\star,S}^{1/3}(T_{\rm nt})}{\langle E_X \rangle\, g_{\star,S}^{1/3}(T_{\rm nr})}.$$

Comparing this to the temperature of a thermal relic,

$$T_{\rm DM}^{\rm thermal} = T_{\rm kd}\left(\frac{1+z}{1+z_{\rm kd}}\right)^2 = \frac{T^2\, g_{\star,S}^{2/3}(T)}{T_{\rm kd}\, g_{\star,S}^{2/3}(T_{\rm kd})},$$

we find that the temperature of the dark matter in this scenario is hotter by the factor

$$\frac{T_{\rm DM}}{T_{\rm DM}^{\rm thermal}} \sim \frac{\langle E_X \rangle^2\, T_{\rm kd}}{9\, T_{\rm nt}^2\, m_X}\left(\frac{g_{\star,S}(T_{\rm kd})}{g_{\star,S}(T_{\rm nt})}\right)^{2/3},$$

which in turn increases the Jeans mass by a factor of $(T_{\rm DM}/T_{\rm DM}^{\rm thermal})^{3/2}$, resulting (up to factors of $g_{\star,S}$) in the following (see Eq. 10.16):

$$M_J \sim 10^{-8}\, M_\odot \times \left(\frac{1\,{\rm GeV}}{m_X}\right)^3 \times \left(\frac{\langle E_X \rangle}{3\, T_{\rm nt}}\right)^3.$$

For the case of $\langle E_X \rangle / T_{\rm nt} = 10^2$ and $m_X = 1\,{\rm GeV}$, the predicted large-scale structure is indistinguishable from that found in the standard case of cold dark matter, with suppression occurring only on very small scales. In the more extreme case of $\langle E_X \rangle / T_{\rm nt} = 10^6$, the growth of structure is suppressed on scales below $M_J \sim 10^8 - 10^9\, M_\odot$, which is roughly equal to the mass of the smallest observed dark matter halos.

Chapter 11

Problem 11.1

We can write this term of the Lagrangian as follows:

$$\mathcal{L} = -m_D \bar{\psi}\psi = -m_D(\bar{\psi}_L + \bar{\psi}_R)(\psi_L + \psi_R) = -m_D(\bar{\psi}_L\psi_L + \bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R + \bar{\psi}_R\psi_R).$$

We can further express the left- and right-handed components of ψ and $\bar{\psi}$ using projection operators:

$$\psi_L = P_L\psi,$$
$$\psi_R = P_R\psi,$$
$$\bar{\psi}_L = \bar{\psi}P_R,$$
$$\bar{\psi}_R = \bar{\psi}P_L,$$

allowing us to see that $\bar{\psi}_L\psi_L = \bar{\psi}P_RP_L\psi = 0$ and $\bar{\psi}_R\psi_R = \bar{\psi}P_LP_R\psi = 0$, leaving us with

$$\mathcal{L} = -m_D(\bar{\psi}_R\psi_L + \bar{\psi}_L\psi_R).$$

Problem 11.3

Starting with the first two terms:

$$\sum_i |U_{\alpha i}|^2|U_{\beta i}|^2 + \sum_{i\neq j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j} = \sum_i\sum_j U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j} = \left|\sum_i U^\star_{\alpha i}U_{\beta i}\right|^2.$$

Since U is unitary, $UU^\dagger$ is equal to the identity matrix, and thus these first two terms can be expressed as $\delta_{\alpha\beta}$.

Moving on to the next term:

$$-2\sum_{i\neq j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right)$$
$$= -2\sum_{i>j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right) - 2\sum_{i<j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right)$$
$$= -2\sum_{i>j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right) - 2\sum_{i>j} U^\star_{\alpha j}U_{\beta j}U_{\alpha i}U^\star_{\beta i}\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right)$$
$$= -2\sum_{i>j}(U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j} + U^\star_{\alpha j}U_{\beta j}U_{\alpha i}U^\star_{\beta i})\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right)$$
$$= -2\sum_{i>j}[U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j} + (U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j})^\star]\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right)$$
$$= -4\sum_{i>j}\mathrm{Re}(U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j})\,\sin^2\left(\frac{(m_i^2-m_j^2)L}{4E}\right).$$

And finally, the last term:

$$-i\sum_{i\neq j} U^\star_{\alpha i}U_{\beta i}U_{\alpha j}U^\star_{\beta j}\,\sin\left(\frac{(m_i^2-m_j^2)L}{2E}\right)$$

$$= -i \sum_{i>j} U^{\star}_{\alpha i} U_{\beta i} U_{\alpha j} U^{\star}_{\beta j} \sin\left(\frac{(m_i^2 - m_j^2)L}{2E}\right) - i \sum_{i<j} U^{\star}_{\alpha i} U_{\beta i} U_{\alpha j} U^{\star}_{\beta j} \sin\left(\frac{(m_i^2 - m_j^2)L}{2E}\right)$$

$$= -i \sum_{i>j} U^{\star}_{\alpha i} U_{\beta i} U_{\alpha j} U^{\star}_{\beta j} \sin\left(\frac{(m_i^2 - m_j^2)L}{2E}\right) + i \sum_{i>j} U^{\star}_{\alpha j} U_{\beta j} U_{\alpha i} U^{\star}_{\beta i} \sin\left(\frac{(m_i^2 - m_j^2)L}{2E}\right)$$

$$= i \sum_{i>j} (-U^{\star}_{\alpha i} U_{\beta i} U_{\alpha j} U^{\star}_{\beta j} + U^{\star}_{\alpha j} U_{\beta j} U_{\alpha i} U^{\star}_{\beta i}) \sin\left(\frac{(m_i^2 - m_j^2)L}{2E}\right)$$

$$= i \sum_{i>j} [-U^{\star}_{\alpha i} U_{\beta i} U_{\alpha j} U^{\star}_{\beta j} + (U^{\star}_{\alpha i} U_{\beta i} U_{\alpha j} U^{\star}_{\beta j})^{\star}] \sin\left(\frac{(m_i^2 - m_j^2)L}{2E}\right)$$

$$= 2 \sum_{i>j} \mathrm{Im}(U^{\star}_{\alpha i} U_{\beta i} U_{\alpha j} U^{\star}_{\beta j}) \sin\left(\frac{(m_i^2 - m_j^2)L}{2E}\right).$$

Combining these terms, we arrive at the final expression given in Eq. 11.22.

Problem 11.4

Using Eq. 11.22, we can write the oscillation probability in this case as follows:

$$P_{\nu_e \to \nu_\tau} \approx -4 \sum_{j=1,2} \mathrm{Re}(U^{\star}_{13} U_{33} U_{1j} U^{\star}_{3j}) \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{4E}\right)$$

$$+ 2 \sum_{j=1,2} \mathrm{Im}(U^{\star}_{13} U_{33} U_{1j} U^{\star}_{3j}) \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{2E}\right)$$

$$= \left[-4\mathrm{Re}(U^{\star}_{13} U_{33} U_{11} U^{\star}_{31}) - 4\mathrm{Re}(U^{\star}_{13} U_{33} U_{12} U^{\star}_{32}) \right] \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{4E}\right)$$

$$+ \left[2\mathrm{Im}(U^{\star}_{13} U_{33} U_{11} U^{\star}_{31}) + 2\mathrm{Im}(U^{\star}_{13} U_{33} U_{12} U^{\star}_{32}) \right] \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{2E}\right)$$

$$= -4 s_{13} c_{13}^2 c_{23}\, \mathrm{Re}\left[c_{12} s_{12} s_{23} e^{i\delta_{\mathrm{CP}}} - c_{12}^2 s_{13} c_{23} - s_{12} c_{12} s_{23} e^{i\delta_{\mathrm{CP}}} - s_{12}^2 s_{13} c_{23} \right] \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{4E}\right)$$

$$+ 2 s_{13} c_{13}^2 c_{23}\, \mathrm{Im}\left[c_{12} s_{12} s_{23} e^{i\delta_{\mathrm{CP}}} - c_{12}^2 s_{13} c_{23} - s_{12} c_{12} s_{23} e^{i\delta_{\mathrm{CP}}} - s_{12}^2 s_{13} c_{23} \right] \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{2E}\right)$$

$$= \left[4 s_{13}^2 c_{13}^2 c_{23}^2 c_{12}^2 + 4 s_{13}^2 c_{13}^2 c_{23}^2 s_{12}^2 \right] \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{4E}\right)$$

$$= 4 s_{13}^2 c_{13}^2 c_{23}^2 \sin\left(\frac{\Delta m^2_{\mathrm{atm}} L}{4E}\right)$$

$$= \sin^2(2\theta_{13}) \cos^2(\theta_{23}) \sin^2\left(\frac{\Delta m^2_{\mathrm{atm}} L}{4E}\right).$$

A similar approach can be used to find the corresponding expressions for $P_{\nu_\mu \to \nu_\tau}$ and $P_{\nu_e \to \nu_\mu}$.

Problem 11.5

In the core of the sun, the matter-induced potential is $V = \sqrt{2} G_F n_e \sim 3.7 \times 10^{-12}$ eV, leading to

$$\begin{aligned}\Delta m_m^2 &= \Delta m_{12}^2 \sqrt{\sin^2 2\theta_{12} + (\cos 2\theta_{12} - 2EV/\Delta m_{12}^2)^2} \\ &\approx 7.5 \times 10^{-5}\,\text{eV}^2 \sqrt{0.31 + [0.38 - 0.099\,(E/\text{MeV})]^2}\end{aligned}$$

and

$$\begin{aligned}\sin^2 2\theta_m &= \frac{\sin^2 2\theta_{12}}{\sin^2 2\theta_{12} + (\cos 2\theta_{12} - 2EV/\Delta m_{12}^2)^2} \\ &\approx \frac{0.31}{0.31 + [0.38 - 0.099\,(E/\text{MeV})]^2},\end{aligned}$$

allowing us to calculate the resulting oscillation probability:

$$P_{\nu_e \to \nu_\mu} = \sin^2 2\theta_m \, \sin^2\left(\frac{\Delta m_m^2 L}{4E}\right),$$

where $L \sim 3 \times 10^7$ cm.

Resonance occurs when $0.38 \approx 0.099\,(E/\text{MeV})$, corresponding to $E \approx 3.8$ MeV. At this energy, the oscillation probability becomes

$$P_{\nu_e \to \nu_\mu} \approx \sin^2\left(\frac{L}{4 \times 10^6\,\text{cm}}\right),$$

and thus oscillations occur many times before the neutrino escapes the sun's core. Note that neutrinos with $E > 3.8$ MeV may also satisfy the resonance condition as they travel through regions of the sun with a lower density, and thus induce a smaller value of V.

Chapter 12

Problem 12.1

At $T \sim 0.1\,M_X$ the energy densities of the Standard Model bath and the X-population are given by:

$$\rho_{\text{SM}} \sim \frac{\pi^2 g_\star\,(0.1\,M_X)^4}{30}, \qquad \rho_X \approx 3 M_X^{5/2}\left(\frac{0.1\,M_X}{2\pi}\right)^{3/2} e^{-10},$$

where I have used $g_X = 3$ for the number of internal degrees of freedom. As the universe further expands, the energy density in the Standard Model bath evolves as $\rho_{\text{SM}} \propto T^4$, while the nonrelativistic bosons dilute according to $\rho_X \propto T^3$. Thus $\rho_{\text{SM}} \approx \rho_X$ when:

$$1 = \frac{\rho_{\text{SM}}}{\rho_X} \approx \frac{\pi^2 g_\star\,(0.1\,M_X)^4}{30}\,\frac{1}{3M_X^{5/2}}\left(\frac{2\pi}{0.1\,M_X}\right)^{3/2} e^{10} \times \frac{(T/0.1M_X)^4}{(T/0.1M_X)^3},$$

which, for $g_\star \approx 106.75$, reduces to $T \approx 8 \times 10^{-6}\,M_X$.

Repeating this for the case where the X population freezes out at $T \sim 10 M_X$, we have $\rho_X \approx 3\pi^2 T^4/30$ at the departure from equilibrium. Until $T \sim M_X$, both energy densities evolve as $\rho_{\rm SM} \propto T^4$, but once the Xs become nonrelativistic, they transition to $\rho_X \propto T^3$. Thus,

$$1 = \frac{\rho_{\rm SM}}{\rho_X} \sim \frac{\pi^2 g_\star \, (10\, M_X)^4}{30} \, \frac{30}{3\pi^2 \, (10 M_X)^4} \times \frac{(T/M_X)^4}{(T/M_X)^3}.$$

Again using $g_\star \approx 106.75$, we find that $\rho_X \gtrsim \rho_{\rm SM}$ starting when $T \sim 0.03\, M_X$.

Problem 12.3

The scattering rate of a fermion for a baryon number violating process mediated by an X is given by

$$\Gamma_S \approx \sigma \, n_f \sim \frac{g_{\rm GUT}^4 T^2}{16\pi^2 M_X^4} \times \frac{3}{4} \frac{\zeta(3) g_f T^3}{\pi^2}.$$

Comparing this to the decay rate, $\Gamma_X \sim g_{\rm GUT}^2 M_X/4\pi$, we have:

$$\frac{\Gamma_S}{\Gamma_X} \sim \frac{3\zeta(3) g_f}{16\pi^3} \frac{g_{\rm GUT}^2 T^5}{M_X^5} \sim 4 \times 10^{-5} \times \left(\frac{g_f}{4}\right)\left(\frac{T}{M_X}\right)^4.$$

Thus during the period of interest, the rates for baryon number violating scattering processes are much lower than those for X decay.

Problem 12.4

The term accounting for the interference between a tree-level diagram and an N-loop diagram for the decay of a heavy GUT boson, analogous to those given in Eq. 12.12, will include $1 + (1 + 2N)$ powers of the couplings (or their complex conjugates). Thus up to ratios of phase space factors and baryon number assignments, we expect such decays to produce no more baryon number than $\epsilon_X \sim 2 g_{\rm GUT}^{2+2N-2}$. For interference with a 2-loop or 3-loop diagram, this could yield a value as large as $\epsilon_X \sim 5 \times 10^{-6}$ or $\epsilon_X \sim 8 \times 10^{-9}$, respectively. Interference with higher order loops cannot produce enough net baryon number to explain the observed asymmetry.

Problem 12.5

Immediately prior to the decays, the energy density of the universe was given by $\rho \approx Mn$, where n is the number density of the decaying particles. After the decays, it is instead given by $\rho = \pi^2 g_\star(T)\, T^4/30$. Setting these equal, we have

$$n \approx \frac{\pi^2 g_\star(T)\, T^4}{30M}.$$

Further expressing $n_B = \epsilon\, n$ and relating the entropy to T and $g_\star$, we have

$$Y_B = \frac{n_B}{s} = \frac{45\,\epsilon\, n}{2\pi^2 g_{\star,S}(T)} \approx \frac{3\,\epsilon\, T}{4M}.$$

Problem 12.6

The Hubble radius contains roughly $N_b \sim (\rho_b/m_p)\, H^{-3} \sim 10^{78}$ baryons. For a weak-scale scattering cross section, $\sigma\, v \sim G_F^2$, and treating the universe as homogeneous, the scattering rate per particle is roughly $\Gamma \sim (\rho_b/m_p)\, \sigma\, v \sim 10^{-33}\ \mathrm{s}^{-1}$. Combining these quantities with the sphaleron suppression factor of $e^{-16\pi^2/g_2^2} \sim 10^{-161}$, we estimate a total rate of sphaleron transitions given by $\sim 10^{-116}\ \mathrm{s}^{-1} \sim 10^{-108}\ \mathrm{yr}^{-1}$. Thus in our universe's current configuration, it would be $\sim 10^{108}$ yr before we should expect a sphaleron transition to take place.

Problem 12.7

The evolution of baryon number (or $B + L$) is dictated by the following differential equation:

$$\frac{dn_B}{dt} = -3Hn_B - \frac{975\, g_2^{10}\, T\, n_B}{2\,(4\pi)^5}.$$

As demonstrated in the chapter text, the sphaleron term is much greater than the Hubble term, so we can neglect the $-3Hn_B$ piece in our estimate. This can now be rewritten as

$$\int_{n_{B,0}}^{n_{B,f}} \frac{dn_B}{n_B} \sim -\frac{975\, g_2^{10}}{2\,(4\pi)^5} \int_{t_0}^{t_f} T\, dt,$$

which reduces to

$$\begin{aligned} \ln\left(\frac{n_{B,f}}{n_{B,0}}\right) &\sim \frac{975\, g_2^{10}}{2\,(4\pi)^5\, M_{\mathrm{Pl}}} \left(\frac{90}{8\pi^3 g_\star}\right)^{0.5} \int_{T_0}^{T_f} \frac{dT}{T^2}, \\ &\sim -\frac{975\, g_2^{10}}{2\,(4\pi)^5\, M_{\mathrm{Pl}}} \left(\frac{90}{8\pi^3 g_\star}\right)^{0.5} \left(\frac{1}{3T_f^3} - \frac{1}{3T_0^3}\right). \end{aligned}$$

Putting in $n_{B,0} \sim T^3$ and $n_{B,f} \sim Y_B \times s \sim Y_B \times T^3$, and solving for T_f, we find that the asymmetry is washed out around $T_f \sim 5$ TeV for $T_0 = 10^{10}$ GeV, and around $T_f \approx 997$ GeV for $T_0 = 1$ TeV.

Problem 12.8

The number of generations comes into this calculation through the condition of Eq. 12.22. Generalizing this for an arbitrary value of N_g, we arrive at the following relationships between the chemical potentials:

$$\mu_{e_R} = \frac{2N_g + 3}{6N_g + 3}\, \mu_{e_L}, \quad \mu_{d_R} = -\frac{6N_g + 1}{6N_g + 3}\, \mu_{e_L}, \quad \mu_{u_R} = \frac{2N_g - 1}{6N_g + 3}\, \mu_{e_L},$$

$$\mu_{u_L} = \mu_{d_L} = -\frac{1}{3}\, \mu_{e_L}, \quad \mu_{\nu_L} = \mu_{e_L}, \quad \mu_H = \frac{4N_g}{6N_g + 3}\, \mu_{e_L}.$$

Notice that one recovers Eq. 12.24 in the case of $N_g = 3$.

Using these relationships, we can write expressions for n_B and n_L:

$$\begin{aligned} n_B &= \left(\frac{\mu_{u_L}T^2}{2} + \frac{\mu_{u_R}T^2}{2} + \frac{\mu_{d_L}T^2}{2} + \frac{\mu_{d_R}T^2}{2}\right) \\ &= \frac{T^2}{2}\left(-\frac{1}{3}\mu_{e_L} + \frac{2N_g-1}{6N_g+3}\mu_{e_L} - \frac{1}{3}\mu_{e_L} - \frac{6N_g+1}{6N_g+3}\mu_{e_L}\right) \\ &= -\frac{4N_g+2}{6N_g+3}\,T^2\,\mu_{e_L}, \end{aligned}$$

$$\begin{aligned} n_L &= 3\times\left(\frac{\mu_{e_L}T^2}{6} + \frac{\mu_{e_R}T^2}{6} + \frac{\mu_{\nu_L}T^2}{6}\right) \\ &= \frac{T^2}{2}\left(\mu_{e_L} + \frac{2N_g+3}{6N_g+3}\,\mu_{e_L} + \mu_{e_L}\right) \\ &= \frac{14N_g+9}{12N_g+6}\,T^2\,\mu_{e_L}. \end{aligned}$$

Taking the ratio of these expression and rewriting in terms of n_{B-L}, we arrive at:

$$n_B = \frac{8N_g+4}{22N_g+13}\times n_{B-L} \quad \text{and} \quad n_L = \left(1-\frac{8N_g+4}{22N_g+13}\right)\times n_{B-L}.$$

Problem 12.11

For the first expression,

$$\begin{aligned} \rho(z) &= \frac{1}{2}\left(\frac{d\phi}{dz}\right)^2 + V(\phi) \\ &= \frac{1}{2}\left[\frac{m^2}{\lambda^{1/2}\sqrt{2}}\cosh^{-2}\left(\frac{z|m|}{\sqrt{2}}\right)\right]^2 + \left[\frac{m^4}{4\lambda^2} - \frac{m^4}{2\lambda}\tanh^2\left(\frac{z|m|}{\sqrt{2}}\right) + \frac{m^4}{4\lambda}\tanh^4\left(\frac{z|m|}{\sqrt{2}}\right)\right] \\ &= \frac{m^4}{4\lambda}\cosh^{-4}\left(\frac{z|m|}{\sqrt{2}}\right) + \frac{m^4}{4\lambda^2}\left[1-\tanh^2\left(\frac{z|m|}{\sqrt{2}}\right)\right]^2 \\ &= \frac{m^4}{2\lambda^2}\cosh^{-4}\left(\frac{z|m|}{\sqrt{2}}\right). \end{aligned}$$

Next, integrating with respect to z, we have

$$\begin{aligned} \eta &= \frac{m^4}{2\lambda}\int_{-\infty}^{\infty}\cosh^{-4}\left(\frac{z|m|}{\sqrt{2}}\right)dz = \frac{m^4}{2\lambda}\int_{-\infty}^{\infty}\left[\tanh^2\left(\frac{z|m|}{\sqrt{2}}\right)-1\right]^2 dz \\ &= \frac{m^4}{2\lambda}\int_{-\infty}^{\infty}\left[\tanh^4\left(\frac{z|m|}{\sqrt{2}}\right) - 2\tanh^2\left(\frac{z|m|}{\sqrt{2}}\right)+1\right]dz \end{aligned}$$

$$
\begin{aligned}
&= \frac{m^4}{2\lambda}\left[-\frac{\sqrt{2}}{3|m|}\tanh^3\left(\frac{z|m|}{\sqrt{2}}\right)\Big|_{-\infty}^{\infty} + \int_{-\infty}^{\infty}\tanh^2\left(\frac{z|m|}{\sqrt{2}}\right)dz \right. \\
&\quad \left. -2\int_{-\infty}^{\infty}\tanh^2\left(\frac{z|m|}{\sqrt{2}}\right)dz + z\Big|_{-\infty}^{\infty}\right] \\
&= \frac{m^4}{2\lambda}\left[-\frac{\sqrt{2}}{3|m|}\tanh^3\left(\frac{z|m|}{\sqrt{2}}\right) - z + \frac{\sqrt{2}}{|m|}\tanh\left(\frac{z|m|}{\sqrt{2}}\right) + z\right]\Big|_{-\infty}^{\infty} \\
&= \frac{|m|^3}{2\lambda}\left[-\frac{\sqrt{2}}{3}(1+1) + \sqrt{2}(1+1)\right] \\
&= \frac{2\sqrt{2}\,|m|^3}{3\lambda}.
\end{aligned}
$$

Problem 12.12

To find the extrema, we set the derivative of the potential equal to zero:

$$0 = \frac{\partial V}{\partial \phi} = \frac{\lambda N}{v}\sin\left(\frac{N\phi}{v}\right),$$

which corresponds to $\phi = 0, \pi v/N, 2\pi v/N, \ldots$. To determine which of these are minima, we evaluate which have a positive second derivative:

$$\frac{\partial^2 V}{\partial \phi^2} = \frac{\lambda N^2}{v^2}\cos\left(\frac{N\phi}{v}\right) > 0,$$

which is satisfied for each of the odd numbered extrema, including $\phi = 0, 2\pi v/N$, $4\pi v/N \ldots$. In order to obtain $V = 0$ at each of these minima, we set $V_0 = 0$.

Next, we show that the function

$$\phi(z) = \frac{4v}{N}\tan^{-1}\left(e^{Bz}\right)$$

is a solution to the equation of motion:

$$\frac{\partial^2 \phi}{\partial z^2} = \frac{dV}{d\phi}.$$

The first and second derivatives of ϕ are given by

$$
\begin{aligned}
\frac{\partial \phi}{\partial z} &= \frac{4v}{N}\frac{Be^{Bz}}{[1+e^{2Bz}]} \\
\frac{\partial^2 \phi}{\partial z^2} &= \frac{4v}{N}\frac{B^2 e^{Bz}}{[1+e^{2Bz}]} - \frac{4v}{N}\frac{2B^2 e^{3Bz}}{[1+e^{2Bz}]^2},
\end{aligned}
$$

while the right-hand side of the equation of motion is

$$\frac{dV}{d\phi} = \frac{\lambda N}{v} \sin\left[4\tan^{-1}\left(e^{Bz}\right)\right].$$

Applying the trigonometric identities $\sin(2x) = 2\sin(x)\cos(x)$ and $\cos(2x) = 1 - 2\sin^2 x$, this can be rewritten as follows:

$$\frac{\lambda N}{v} \sin\left[4\tan^{-1}\left(e^{Bz}\right)\right] = \frac{2\lambda N}{v} \sin\left[2\tan^{-1}\left(e^{Bz}\right)\right] \cos\left[2\tan^{-1}\left(e^{Bz}\right)\right]$$
$$= \frac{4\lambda N}{v} \sin\left[\tan^{-1}\left(e^{Bz}\right)\right] \cos\left[\tan^{-1}\left(e^{Bz}\right)\right]\left[1 - 2\sin^2\left[\tan^{-1}\left(e^{Bz}\right)\right]\right].$$

Next, applying the trigonometric identities $\sin(\tan^{-1})x = x/\sqrt{1+x^2}$ and $\cos(\tan^{-1})x = 1/\sqrt{1+x^2}$, we arrive at the following:

$$\frac{\lambda N}{v} \sin\left[4\tan^{-1}\left(e^{Bz}\right)\right] = \frac{4\lambda N}{v} \frac{e^{Bz}}{[1+e^{2Bz}]}\left[1 - \frac{2e^{2Bz}}{[1+e^{2Bz}]}\right]$$
$$= \frac{4\lambda N}{v} \frac{e^{Bz}}{[1+e^{2Bz}]} - \frac{4\lambda N}{v} \frac{e^{3Bz}}{[1+e^{2Bz}]^2}.$$

The equation of motion can now be written as:

$$\frac{4v}{N} \frac{B^2 e^{Bz}}{[1+e^{2Bz}]} - \frac{4v}{N} \frac{2B^2 e^{3Bz}}{[1+e^{2Bz}]^2} = \frac{4\lambda N}{v} \frac{e^{Bz}}{[1+e^{2Bz}]} - \frac{4\lambda N}{v} \frac{e^{3Bz}}{[1+e^{2Bz}]^2},$$

which is satisfied for $B^2 = \lambda N^2/v^2$.

To find the energy density per area of the resulting domain walls, we integrate the energy density with respect to z, from the value corresponding to one minimum to that of an adjacent minimum. For example, for $\phi = 0$ and $\phi = 2\pi v/N$, we can use the equation of motion to show that the corresponding limits of integration are $z = -\infty$ and $z = \infty$.

The energy density per area is then given by

$$\eta = \int_{-\infty}^{\infty} \rho(z)dz = \int_{-\infty}^{\infty} \left[\frac{1}{2}\left(\frac{d\phi}{dz}\right)^2 + V(\phi)\right]dz$$
$$= \int_{-\infty}^{\infty} \left[\frac{1}{2}\left(\frac{4\lambda^{1/2} e^{\lambda^{1/2}Nz/v}}{[1+e^{2\lambda^{1/2}Nz/v}]}\right)^2 + \lambda\left[1 - \cos[4\tan^{-1}(e^{\lambda^{1/2}Nz/v})]\right]\right]dz.$$

Using the same trigonometric identities we applied earlier in this problem, we can rewrite the cosine in this expression as follows:

$$\cos[4\tan^{-1}(e^{\lambda^{1/2}Nz/v})] = 1 - 2\sin^2[8\tan^{-1}(e^{\lambda^{1/2}Nz/v})]$$
$$= 1 - 2\left[2\sin[\tan^{-1}(e^{\lambda^{1/2}Nz/v})]\cos[\tan^{-1}(e^{\lambda^{1/2}Nz/v})]\right]^2$$

$$= 1 - 8\sin^2[\tan^{-1}(e^{\lambda^{1/2}Nz/v})]\cos^2[\tan^{-1}(e^{\lambda^{1/2}Nz/v})]$$
$$= 1 - \frac{8e^{2\lambda^{1/2}Nz/v}}{(1+e^{2\lambda^{1/2}Nz/v})^2}.$$

Inserting this into the expression for η, we now have

$$\eta = \int_{-\infty}^{\infty}\left[\frac{1}{2}\left(\frac{4\lambda^{1/2}e^{\lambda^{1/2}Nz/v}}{[1+e^{2\lambda^{1/2}Nz/v}]}\right)^2 + \lambda\left[1-1+\frac{8e^{2\lambda^{1/2}Nz/v}}{(1+e^{2\lambda^{1/2}Nz/v})^2}\right]\right]dz$$
$$= \int_{-\infty}^{\infty}\frac{16\lambda e^{2\lambda^{1/2}Nz/v}}{[1+e^{2\lambda^{1/2}Nz/v}]^2}dz = 16\lambda\left[\frac{v}{2N\lambda^{1/2}} - 0\right] = \frac{8v\lambda^{1/2}}{N}.$$

Problem 12.14

The Boltzmann equation describing the evolution of the number density of monopoles is given by

$$\frac{dn_M}{dt} = -3Hn_M - \langle\sigma v\rangle n_M^2.$$

The ratio of the annihilation and Hubble terms is given by

$$\frac{\langle\sigma v\rangle n_M^2}{3Hn_M} = \frac{(n_M/s)s}{16\pi^2 Tm_M}\frac{1}{3}\left(\frac{3M_{\rm Pl}^2}{8\pi\rho}\right)^{1/2}$$
$$= \frac{(5\pi g_\star)^{1/2}}{45}\frac{M_{\rm Pl}}{m_M}\frac{n_M}{s}$$
$$\approx 2\times 10^{-9}\times\left(\frac{g_\star(T)}{200}\right)^{1/2}.$$

Thus annihilations do not appreciably reduce the abundance of monopoles.

Chapter 13

Problem 13.2

As dark energy became the dominant form of energy in our universe, the evolution of curvature evolved from $|\Omega_\kappa| \propto a$ during matter domination, to $|\Omega_\kappa| \propto a^{-2}$ during dark radiation domination. To find out at what scale factor the magnitude of curvature began to decrease, we set the derivative of $|\Omega_\kappa|$ with respect to a to zero:

$$0 = \frac{d(a^{-2}\rho^{-1})}{da} = \frac{d(a^{-2}[\Omega_M a^{-3} + \Omega_\Lambda]^{-1})}{da},$$

whose solution is $a = (\Omega_M/2\Omega_\Lambda)^{1/3} \approx 0.61$. In other words, our universe's curvature reached a maximum in magnitude when $a \approx 0.61$ and has since been becoming more flat.

Problem 13.6

The Hubble rate is given by $H = (8\pi\rho/3M_{\rm Pl}^2)^{1/2}$, where $\rho = \dot{\phi}^2/2 + V$. Taking the time derivative, we have

$$\begin{aligned}
\dot{H} &= \left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2} \frac{d}{dt}\left[\left(\frac{\dot{\phi}^2}{2} + V\right)^{1/2}\right] \\
&= \left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2} \frac{1}{2}\left(\frac{\dot{\phi}^2}{2} + V\right)^{-1/2}\left[\dot{\phi}\,\ddot{\phi} + \frac{dV}{dt}\right] \\
&= \left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2} \frac{1}{2}\left(\frac{\dot{\phi}^2}{2} + V\right)^{-1/2}\left[\dot{\phi}\left(-3H\dot{\phi} - \frac{dV}{d\phi}\right) + \frac{dV}{dt}\right] \\
&= \left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2} \frac{1}{2}\left(\frac{8\pi}{3M_{\rm Pl}^2H^2}\right)^{1/2}\left[\dot{\phi}\left(-3H\dot{\phi} - \frac{dV}{d\phi}\right) + \frac{dV}{dt}\right] \\
&= \frac{4\pi}{3M_{\rm Pl}^2H}\left[-3H\dot{\phi}^2 - \dot{\phi}\frac{dV}{d\phi} + \frac{dV}{dt}\right] \\
&= -\frac{4\pi\dot{\phi}^2}{M_{\rm Pl}^2}.
\end{aligned}$$

Problem 13.7

Starting with Eq. 13.14,

$$\frac{\ddot{a}}{a} = H^2\,(1-\epsilon),$$

the solution to this equation is given by $a \propto e^{tH(1-\epsilon)^{1/2}}$. This corresponds to one e-fold of expansion per $\Delta t = 1/(H(1-\epsilon)^{1/2})$. From this, we can write

$$\frac{dN}{dt} = H\,(1-\epsilon)^{1/2} \approx H,$$

where in the last step I have assumed that $\epsilon \ll 1$. Next, we can write

$$\frac{dH}{dN} = \frac{dH}{dt}\frac{dt}{dN} \approx \frac{\dot{H}}{H},$$

and thus

$$\epsilon = -\frac{\dot{H}}{H^2} \approx -\frac{1}{H}\frac{dH}{dN} = -\frac{d\ln H}{dN}.$$

Problem 13.8

$$-\frac{d\ln|dN/d\phi|}{dN} = -\frac{1}{dH/d\phi}\frac{d(dH/d\phi)}{dN} = -\frac{1}{dH/d\phi}\frac{d^2H}{d\phi^2}\frac{d\phi}{dN}$$

$$= -\frac{1}{dH/d\phi}\frac{d^2H}{d\phi^2}\frac{d\phi}{dt}\frac{dt}{dN} = -\frac{1}{dH/d\phi}\frac{d^2H}{d\phi^2}\left(-\frac{M_{\rm Pl}^2}{4\pi}\frac{dH}{d\phi}\right)\left(\frac{1}{H}\right)$$
$$= \frac{M_{\rm Pl}^2}{4\pi H}\frac{d^2H}{d\phi^2}.$$

Problem 13.9

For $V \gg \dot{\phi}^2/2$, we can write $H^2 \approx 8\pi V/3M_{\rm Pl}^2$. Also, for small $\ddot{\phi}$, we can write $dV/d\phi \approx -3H\dot{\phi}$. Together, we can express ϵ_v as follows:

$$\epsilon_v = \frac{M_{\rm Pl}^2}{16\pi}\left(\frac{1}{V}\frac{dV}{d\phi}\right)^2 \approx \frac{M_{\rm Pl}^2}{16\pi}\left(\frac{8\pi}{3M_{\rm Pl}^2H^2}\right)^2\left(-3H\dot{\phi}\right)^2 = \frac{4\pi\dot{\phi}^2}{M_{\rm Pl}^2H^2} = \epsilon.$$

Problem 13.10

For $V \gg \dot{\phi}^2/2$, we can write $H^2 \approx 8\pi V/3M_{\rm Pl}^2$. Using this, we can evaluate $d^2H/d\phi^2$:

$$\frac{d}{d\phi}\left[\frac{dH}{d\phi}\right] = \frac{d}{d\phi}\left[\left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2}\frac{1}{2V^{1/2}}\frac{dV}{d\phi}\right] = \left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2}\frac{1}{2}\left[-\frac{1}{2V^{3/2}}\left(\frac{dV}{d\phi}\right)^2 + \frac{1}{V^{1/2}}\frac{d^2V}{d\phi^2}\right].$$

Inserting this into the definition of η, we have

$$\eta \equiv \frac{M_{\rm Pl}^2}{4\pi H}\frac{d^2H}{d\phi^2} \approx \frac{M_{\rm Pl}^2}{4\pi}\left(\frac{3M_{\rm Pl}^2}{8\pi V}\right)^{1/2}\left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2}\frac{1}{2}\left[-\frac{1}{2V^{3/2}}\left(\frac{dV}{d\phi}\right)^2 + \frac{1}{V^{1/2}}\frac{d^2V}{d\phi^2}\right]$$
$$\approx \frac{M_{\rm Pl}^2}{8\pi\ V^{1/2}}\left[-\frac{1}{2V^{3/2}}\left(\frac{dV}{d\phi}\right)^2 + \frac{1}{V^{1/2}}\frac{d^2V}{d\phi^2}\right]$$
$$= -\frac{M_{\rm Pl}^2}{16\pi V^2}\left(\frac{dV}{d\phi}\right)^2 + \frac{M_{\rm Pl}^2}{8\pi V}\frac{d^2V}{d\phi^2} = -\epsilon_v + \eta_v.$$

Problem 13.11

From Eqs. 13.24 and 13.25, we have

$$A_s \approx \frac{8V}{3\epsilon_v M_{\rm Pl}^4}.$$

In the $V = m_\phi^2\phi^2/2$ model, $\epsilon_v = M_{\rm Pl}^2/4\pi\phi^2$, so

$$A_s \approx \frac{8}{3M_{\rm Pl}^4}\frac{m_\phi^2\phi^2}{2}\frac{4\pi\phi^2}{M_{\rm Pl}^2} \approx \frac{16\pi m_\phi^2\phi^4}{3M_{\rm Pl}^6} \approx 10^{-35}\ {\rm GeV}^{-2}\ m_\phi^2,$$

where in the last step I used $\phi = \phi_{\rm CMB} \approx 3.1M_{\rm Pl}$, which yields $N_{\rm CMB} \approx 60$ in this model. For the measured value of $A_s \approx 2.1 \times 10^{-9}$, this requires $m_\phi \approx 1.4 \times 10^{13}$ GeV.

Problem 13.12

For the potential $V = V_0\phi^4$, the first, second, and third derivatives are given by $dV/d\phi = 4V_0\phi^3$, $d^2V/d\phi^2 = 12V_0\phi^2$, and $d^3V/d\phi^3 = 24V_0\phi$. This yields the following slow-roll

parameters:

$$\epsilon_v = \frac{M_{\rm Pl}^2}{16\pi}\left(\frac{4V_0\phi^3}{V_0\phi^4}\right)^2 = \frac{M_{\rm Pl}^2}{\pi\phi^2},$$

$$\eta_v = \frac{M_{\rm Pl}^2}{8\pi}\frac{12V_0\phi^2}{V_0\phi^4} = \frac{3M_{\rm Pl}^2}{2\pi\phi^2}.$$

Setting $\epsilon_v = 1$, we find that $\phi_{\rm end} = M_{\rm Pl}/\sqrt{\pi}$. We then find $\phi_{\rm CMB}$ by solving the following:

$$N_{\rm CMB} = -\int_{\phi_{\rm CMB}}^{\phi_{\rm end}}\left(\frac{4\pi}{M_{\rm Pl}^2\epsilon}\right)^{1/2} d\phi = -\frac{2\pi}{M_{\rm Pl}^2}\int_{\phi_{\rm CMB}}^{M_{\rm Pl}/\sqrt{\pi}} \phi\, d\phi,$$

which yields $\phi_{\rm CMB}^2 = M_{\rm Pl}^2(N_{\rm CMB}+1)/\pi$. When evaluated at $\phi = \phi_{\rm CMB}$, the slow-roll parameters are $\epsilon_v = 1/(N_{\rm CMB}+1)$ and $\eta_v = 3/2(N_{\rm CMB}+1)$. These result in the following observable parameters:

$$n_s = 1 - 6\epsilon_v + 2\eta = 1 - \frac{3}{N_{\rm CMB}+1} \approx 0.941 - 0.951,$$

$$\frac{dn_s}{d\ln k} = -8\epsilon_v(3\epsilon_v - 2\eta_v) - 2\xi^2 = -\left(\frac{3}{2(N_{\rm CMB}+1)^2}\right) \approx 0.00040 - 0.00058,$$

$$r = 16\epsilon_v = \frac{16}{N_{\rm CMB}+1} \approx 0.26 - 0.31.$$

The tensor-to-scalar ratio predicted by this model is observationally excluded.

Problem 13.13

Repeating the exercise described in Sec. 13.6.1 for the case of this more general potential, one arrives at $r = 16n/(4N_{\rm CMB}+n) \approx 4n/N_{\rm CMB}$.

Problem 13.14

From the first and second derivatives of the potential, we calculate the slow-roll parameters that result in this model:

$$\epsilon_v = \frac{\phi^6 M_{\rm Pl}^2}{\pi\phi_0^8\,[1-(\phi/\phi_0)^4]^2},$$

$$\eta_v = -\frac{3\phi^2 M_{\rm Pl}^2}{2\pi\phi_0^4\,[1-(\phi/\phi_0)^4]}.$$

Next, we can express the tensor-to-scalar ratio, r, in terms of n_s and η_v/ϵ_v:

$$n_s = 1 - 6\epsilon_v + 2\frac{\eta_v}{\epsilon_v}\epsilon_v = 1 + \epsilon_v\left[-6 + \frac{2\eta_v}{\epsilon_v}\right] = 1 + \frac{r}{16}\left[-6 + \frac{2\eta_v}{\epsilon_v}\right],$$

and thus

$$r = \frac{8\,(1-n_s)}{\left[3 - \dfrac{\eta_v}{\epsilon_v}\right]}.$$

In this model, the ratio of the slow-roll parameters is given by

$$\frac{\eta_v}{\epsilon_v} = -\frac{3(1-x^4)}{2x^4},$$

where $x \equiv \phi_{\rm CMB}/\phi_0$. Together, this yields

$$r = \frac{8\,(1-n_s)}{\left[3 + \dfrac{3[1-x^4]}{2x^4}\right]} = \frac{8}{3}\,\frac{(1-n_s)}{\left[1 + \dfrac{(1-x^4)}{2x^4}\right]}.$$

Problem 13.15

After calculating the first and second derivatives of the potential, we find that the slow-roll parameters are given by the following:

$$\epsilon_v = \frac{\sin^2(\phi/\phi_0)}{16\pi G\phi_0^2[1+\cos(\phi/\phi_0)]^2},$$

$$\eta_v = \frac{-\cos^2(\phi/\phi_0)}{8\pi G\phi_0^2[1+\cos(\phi/\phi_0)]}.$$

Setting $\epsilon_v = 1$ and $\phi_0 = 2M_{\rm Pl}$, we find $\phi_{\rm end} \approx 6.56\,M_{\rm Pl}$. Then using this expression for ϵ_v in Eq. 13.22, we find that $\phi_{\rm CMB} \approx (9.00 - 9.23)\,M_{\rm Pl}$ for $N_{\rm CMB} = 50{-}60$. Using this range of values for $\phi_{\rm CMB}$, we find $n_s \approx 0.960{-}0.966$ and $r \approx 0.097{-}0.122$.

Problem 13.16

In this model, the conditions for eternal inflation can be written as

$$\frac{V^{3/2}}{|dV/d\phi|} = \frac{V_0^{1/2}\phi^3}{4} \gtrsim 0.052\,M_{\rm Pl}^3.$$

From the measured value of A_s, we can determine V_0:

$$A_s = \frac{8V}{3\epsilon_v M_{\rm Pl}^4} = \frac{8V}{3M_{\rm Pl}^4}\left(\frac{16\pi V^2}{M_{\rm Pl}^2\,(dV/d\phi)^2}\right) = \frac{128\pi V_0^3\phi^{12}}{3M_{\rm Pl}^6\,(4V_0\phi^3)^2} = \frac{8\pi V_o\phi_{\rm CMB}^6}{3M_{\rm Pl}^6}.$$

Combining these expressions, we have

$$\frac{V_0^{1/2}\phi^3}{4} = \frac{1}{4}\left(\frac{3M_{\rm Pl}^6 A_s}{8\pi}\right)^{1/2}\left(\frac{\phi}{\phi_{\rm CMB}}\right)^3 \gtrsim 0.052\,M_{\rm Pl}^3,$$

which for $A_s = 2.1 \times 10^{-9}$ yields $\phi \gtrsim 24\phi_{\rm CMB}$. In Problem 13.12, we found that $\phi_{\rm CMB} = M_{\rm Pl}[(N_{\rm CMB}+1)/\pi]^{1/2} \approx 4.4\, M_{\rm Pl}$ in this model. Thus eternal inflation proceeds if $\phi \gtrsim 10^2\, M_{\rm Pl}$, corresponding to an energy density given by

$$V \gtrsim V_0 \times (10^2\, M_{\rm Pl})^4 = \frac{3A_s M_{\rm Pl}^6 (10^2\, M_{\rm Pl})^4}{8\pi\phi_{\rm CMB}^6} \approx (0.04\, M_{\rm Pl})^4.$$

In this model of inflation, if the energy density was ever greater than $V \gtrsim (0.04\, M_{\rm Pl})^4$, then the total inflating volume would increase monotonically without limit.

Chapter 14

Problem 14.4

For this potential, the equation of motion is given by

$$\ddot{\phi} + \frac{3}{2t}\dot{\phi} - \alpha V_0 \phi^{-\alpha-1} = 0.$$

Taking $\phi = \phi_0 t^\beta$, the equation of motion becomes

$$\phi_0\beta(\beta-1)t^{\beta-2} + \frac{3}{2t}\phi_0\beta t^{\beta-2} - \alpha V_0\,(\phi_0 t^\beta)^{-\alpha-1} = 0.$$

Setting the time exponents equal, we find $\beta = 2/(2+\alpha)$. Using this expression for β and solving for ϕ_0, we find the following solution:

$$\phi = \left(\frac{\alpha\,(2+\alpha)^2\,V_0}{6+\alpha}\right)^{1/(\alpha+2)} t^{2/(2+\alpha)}.$$

Problem 14.5

A small perturbation to the solution given in Eq. 14.2, $\delta\phi$, will evolve as follows:

$$\delta\ddot{\phi} + \frac{3}{2t}\delta\dot{\phi} + \frac{dV}{d\phi}(\phi+\delta\phi) - \frac{dV}{d\phi}(\phi) = 0.$$

Focusing first on the last two terms, we can write

$$\frac{dV}{d\phi}(\phi+\delta\phi) - \frac{dV}{d\phi}(\phi) = -2V_0\Big[(\phi+\delta\phi)^{-3} - \phi^{-3}\Big] \approx \frac{6V_0}{\phi^4}\delta\phi,$$

where in the last step I have performed a Taylor series expansion of $(\phi+\delta\phi)^{-3}$. Taking $\delta\phi = At^B$, the equation of motion,

$$\delta\ddot{\phi} + \frac{3}{2t}\delta\dot{\phi} + \frac{6V_0}{\phi^4}\delta\phi = 0,$$

can be written as

$$AB(B-1)\,t^{B-2} + \frac{3}{2t}AB\,t^{B-1} + \frac{3}{2t^2}\,A\,t^{B} = 0.$$

Solving this for B by applying the quadratic formula, we find $B = -1/4 \pm \sqrt{23}\,i/4$. Since the real part of B is negative, the perturbations from Eq. 14.2 decay with time, $\delta\phi \propto t^{-1/4}$, demonstrating that this is, in fact, a tracker solution.

Problem 14.6

Using $\phi = (3V_0 M_{\rm Pl}^2/2\pi)^{1/6}\,t^{1/3}$, the kinetic energy and potential energy are given by

$$\frac{\dot\phi^2}{2} = \frac{1}{18}\left(\frac{3V_0 M_{\rm Pl}^2}{2\pi}\right)^{1/3} t^{-4/3}$$

and

$$V = V_0\left(\frac{2\pi}{3V_0 M_{\rm Pl}^2}\right)^{1/3} t^{-2/3}.$$

Thus the kinetic energy falls more rapidly than the potential energy, and $\phi = (3V_0 M_{\rm Pl}^2/2\pi)^{1/6}\,t^{1/3}$ is the asymptotic solution to the equation of motion.

Problem 14.7

The equivalent of Eq. 14.4 in this case is given by

$$3\left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2}\left(\frac{V_0}{\phi^\alpha}\right)^{1/2}\dot\phi - \frac{\alpha V_0}{\phi^{\alpha+1}} \approx 0,$$

whose solution is

$$\phi = \left(\frac{V_0^{1/2} M_{\rm Pl}\alpha\,(2+\alpha/2)}{(24\pi)^{1/2}}\right)^{1/(2+\alpha/2)} t^{1/(2+\alpha/2)}.$$

Using $\rho_{\rm tot} \approx \rho_\phi \approx V$, we can find that

$$H \approx \left(\frac{8\pi}{3M_{\rm Pl}^2}\,\frac{V_0}{\phi^\alpha}\right)^{1/2} = \left(\frac{8^{2+\alpha}\pi^{2+\alpha}V_0^2}{9\alpha^\alpha(2+\alpha/2)^\alpha M_{\rm Pl}^{4+2\alpha}}\right)^{1/(4+\alpha)} t^{-\alpha/(4+\alpha)}.$$

The scale factor in this case evolves as

$$a \propto e^{(t/\tau)^{4/(4+\alpha)}},$$

where

$$\tau = \left(\frac{8^{2+\alpha}\pi^{2+\alpha}V_0^2}{9\alpha^\alpha(2+\alpha/2)^\alpha M_{\rm Pl}^{4+2\alpha}}\right)^{-1/4}\left(\frac{4+\alpha}{4}\right)^{-(4+\alpha)/4}.$$

Problem 14.8

In the slow-roll regime, the equation of motion is given by

$$3H\dot{\phi} \approx -\frac{dV}{d\phi},$$

$$3\left(\frac{8\pi\rho}{3M_{\rm Pl}^2}\right)^{1/2}\dot{\phi} \approx \frac{V_0}{\phi_0}\, e^{-\phi/\phi_0}.$$

During radiation domination, $\rho_{\rm tot} \approx 3M_{\rm Pl}^2/(32\pi t^2)$, and thus

$$\frac{3}{2t}\,\dot{\phi} \approx \frac{V_0}{\phi_0}\, e^{-\phi/\phi_0}.$$

The solution in this case takes the form of $\phi = A \ln Bt$, for which the equation of motion becomes

$$\frac{3}{2t}\,\frac{A}{t} \approx \frac{V_0}{\phi_0 B^{A/\phi_0} t^{A/\phi_0}}.$$

This implies that $A = 2\phi_0$ and $B = (V_0/3\phi_0^2)^{1/2}$, and thus that $\phi = 2\phi_0 \ln[\sqrt{V_0/3\phi_0^2}\, t]$.

Repeating this exercise for the case of matter domination, $\rho_{\rm tot} \approx M_{\rm Pl}^2/(6\pi t^2)$, we find that the solution is given by $\phi = 2\phi_0 \ln[V_0^{1/2} t/2\phi_0]$. During the scalar field-dominated era, $\rho_{\rm tot} \approx V$, and thus

$$3\left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2} V_0^{1/2}\, e^{-\phi/2\phi_0}\,\dot{\phi} \approx \frac{V_0}{\phi_0}\, e^{-\phi/\phi_0},$$

whose solution is $\phi = 2\phi_0 \ln[t\, V_0^{1/2} M_{\rm Pl}/4\sqrt{6\pi}\phi_0^2]$.

Using these expressions for ϕ to calculate the energy density of the scalar field, $\rho_\phi \approx V = V_0 e^{-\phi/\phi_0}$, we find that during the radiation- and matter-dominated eras, $\rho_\phi = 3\phi_0^2/t^2$ and $\rho_\phi = 4\phi_0^2/t^2$, respectively. This scaling with t^{-2} demonstrates that this potential does, in fact, lead to a tracking solution, with $\rho_\phi \propto \rho_{\rm tot}$.

Problem 14.9

From the first Friedmann equation, we have

$$\left(\frac{1}{a}\frac{da}{dt}\right)^2 = \frac{8\pi}{3M_{\rm Pl}^2}\left(\rho_\Lambda + \rho_{M,0}\, a^{-3}\right),$$

which can be rewritten as

$$\int_0^{a_{\rm Eq}} \frac{da}{a\,(\rho_\Lambda + \rho_{M,0}\, a^{-3})^{1/2}} = \left(\frac{8\pi}{3M_{\rm Pl}^2}\right)^{1/2} \int_0^{t_{\rm Eq}} dt,$$

where $a_{\rm Eq} = (\rho_{M,0}/\rho_\Lambda)^{1/3}$. If we make the approximation that ρ_Λ in the denominator of the left side is small during matter domination, this reduces to $t_{\rm Eq} \sim (M_{\rm Pl}^2/6\pi\rho_\Lambda)^{1/2}$. For the value of ρ_Λ measured in our universe, this yields $t_{\rm Eq} \sim 11.5$ Gyr, which is not far off from the

correct value of 10.1 Gyr. If we perform the integral numerically without this approximation, we obtain the correct result.

Problem 14.10

Under the stated assumptions, the number of stars that form in a given universe is proportional to $t_{\rm Eq} - t_0$, where $t_{\rm Eq}$ is the time at which $\rho_\Lambda = \rho_M$. From Problem 14.9, we know that $t_{\rm Eq} \sim (M_{\rm Pl}^2/6\pi\rho_\Lambda)^{1/2}$. We can use this to find the value of ρ_Λ for which the median star forms:

$$\frac{1}{2} = \frac{\int_0^{\rho_\Lambda^{\rm med}} [(M_{\rm Pl}^2/6\pi\rho_\Lambda)^{1/2} - t_0]\, d\rho_\Lambda}{\int_0^{\rho_\Lambda^{\rm max}} [(M_{\rm Pl}^2/6\pi\rho_\Lambda)^{1/2} - t_0]\, d\rho_\Lambda},$$

where $\rho_\Lambda^{\rm max} = M_{\rm Pl}^2/(6\pi t_0^2)$ is the maximum energy density of the vacuum for which there was any star formation. Carrying out these integrals, we arrive at

$$\frac{M_{\rm Pl}^2}{12\pi t_0} = \frac{2M_{\rm Pl}\,(\rho_\Lambda^{\rm med})^{1/2}}{(6\pi)^{1/2}} - t_0\rho_\Lambda^{\rm med}.$$

Using the quadratic formula, we obtain

$$\rho_\Lambda^{\rm med} = \left(1 - \frac{1}{2^{1/2}}\right)^2 \frac{M_{\rm Pl}^2}{6\pi t_0^2} \approx 0.09\,\rho_\Lambda^{\rm max}.$$

Chapter 15

Problem 15.1

Solving the energy loss equation, it takes the following time for a 1 TeV electron to cool to 100 GeV:

$$t_{\rm max} = \left(\frac{1}{100\,{\rm GeV}} - \frac{1}{1,000\,{\rm GeV}}\right) 10^{16}{\rm GeV}\cdot{\rm s} = 9\times10^{13}\ {\rm s}.$$

For electrons that were injected a time $t_{\rm inj}$ ago, their spatial distribution will be given by $n \propto e^{-d^2/L_{\rm dif}^2}$, where $L_{\rm dif} = 2\sqrt{Dt_{\rm inj}}$ and d is the distance from the source. Normalizing the integral of this distribution to be equal to the injection rate, $R = 10^{40}\ {\rm s}^{-1}$, we have

$$R = \int \frac{dn}{dt_{\rm inj}}\, dV \longrightarrow \frac{dn}{dt_{\rm inj}} = \frac{R}{\pi^{3/2}L_{\rm dif}^3}\, e^{-d^2/L_{\rm dif}^2}.$$

Integrating this from 0 to $t_{\rm max}$, we get the total number density of electrons above 100 GeV:

$$n = \int_0^{t_{\rm max}} \frac{dn}{dt_{\rm inj}}\, dt_{\rm inj} = \int_0^{t_{\rm max}} \frac{R}{\pi^{3/2}L_{\rm dif}^3}\, e^{-d^2/L_{\rm dif}^2}\, dt_{\rm inj} = \int_0^{t_{\rm max}} \frac{R\, e^{-d^2/(4Dt_{\rm inj})}\, dt_{\rm inj}}{8\pi^{3/2}D^{3/2}t_{\rm inf}^{3/2}}.$$

After making the substitution $u = t^{-1/2}$, we perform this integral and arrive at

$$n = -\int_0^{t_{\rm max}} \frac{R\, e^{-d^2u^2/(4D)}\, du}{4\pi^{3/2}D^{3/2}} = \frac{R}{4\pi^{3/2}D^{3/2}}\frac{\pi^{1/2}D^{1/2}}{d}\,{\rm erf}\left(\frac{d}{2D^{1/2}t_{\rm max}^{1/2}}\right)$$
$$= \frac{R}{4\pi Dd}\,{\rm erf}\left(\frac{d}{2D^{1/2}t_{\rm max}^{1/2}}\right) \approx 1.0\times10^{-13}\,{\rm cm}^{-3}.$$

Problem 15.4

In this case, every collision results in $\Delta E/E \approx 2v_{\rm shock}v_{\rm CR} + 2v_{\rm shock}^2 \approx 2v_{\rm shock}$. Repeating the calculation for Fermi acceleration, but using $dE/dt = 2v_{\rm shock}RE$, we find that this scenario should yield a spectrum of particles of the form $N(E) \propto E^{-1-u/2v_{\rm shock}}$. Next, taking $v_{\rm shock} = 3u/4$ (as discussed in Sec. 15.3.2), we obtain a spectrum of the form $N(E) \propto E^{-5/3}$.

Problem 15.5

Inserting these numbers (as well as $Z = 1$ and $u = c$) into Eq. 15.16, we arrive at $E_{\rm max} \approx$ 11.6 TeV, which is only modestly higher than the value achieved at the LHC, $E = 6.8$ TeV.

Problem 15.6

Over a distance d, a proton will pass through $d/2L_{\rm coh}$ domains with $B = 3\,\mu$G and another $d/2L_{\rm coh}$ domains with $B = 30\,\mu$G. In propagating across the high-field domains, this leads to the following degree of net deflection:

$$\theta \sim \left(\frac{L_{\rm coh}}{R_L(B=30\,\mu{\rm G})}\right)\left(\frac{d}{2L_{\rm coh}}\right)^{1/2}$$
$$\sim 2.5^\circ \times \left(\frac{10^{11}\,{\rm GeV}}{E}\right)\left(\frac{d}{1\,{\rm kpc}}\right)^{1/2},$$

while travel through the low-field domains induces a smaller net deflection of $\theta \sim 0.25^\circ$. Together, this yields $\theta \sim [(2.5^\circ)^2 + (0.25^\circ)^2]^{1/2} \sim 2.5^\circ$.

Chapter 16

Problem 16.2

The CMB is no different in the galaxy than elsewhere in the low-redshift universe, and from Figure 16.2 we see that the mean free path for scattering with the CMB is at its minimum of $\lambda \approx 7$ kpc at an energy of about $E_\gamma \approx 2$ PeV. The blackbody temperature of starlight is approximately 5,000 K, which is greater than that of the CMB by a factor of approximately 2,000, reducing the threshold for pair production to $E_\gamma \sim$ TeV. The number density of starlight photons is much lower than that of the CMB, however, with $n_{\rm star} \approx \rho_{\rm star}/(2.7\,T_{\rm star}) \approx 0.26\,{\rm cm}^{-3}$. Thus the minimum mean free path for scattering with starlight is approximately $\lambda \sim 7\,{\rm kpc} \times (0.26/410.7) \approx 11$ Mpc. A 2 PeV photon traveling 1 kpc in our local galaxy would thus have a probability of scattering that is equal to $P \approx 1 - e^{-1\,{\rm kpc}/7\,{\rm kpc}} \approx 13\%$. At lower energies, scattering with starlight dominates, but with a scattering probability at $E_\gamma = 1$ TeV of only $P \approx 1 - e^{-1\,{\rm kpc}/11\,{\rm Mpc}} \approx 0.01\%$.

Problem 16.3

At energies between $E_\gamma \sim 10^6 - 10^9$ GeV, the mean free path for pair production is negligible ($\lambda <$ Mpc) and the resulting electrons and positrons undergo Inverse Compton scattering in the Klein-Nishina regime ($E_e \gg m_e^2/\epsilon_{\rm CMB}$). As a result, over approximately $\log 10^3 / \log 2 = 10$ cycles of pair production and inverse Compton scattering, the initial EeV photon will be transformed into $\sim 10^3$ photons, each with an energy of around $E_\gamma \sim 1$ PeV. At this point, these photons are still within 1 Mpc or so of their source.

At energies below 1 PeV, inverse Compton scattering with the CMB no longer occurs in the Klein-Nishina limit, and the resulting photons are lower in energy than their parent electrons by a factor of roughly $\langle E_\gamma / E_e \rangle \sim (E_e/{\rm PeV})$ (see Eq. 5.19). For photons less energetic than 100 TeV, however, this scattering is dominated by infrared photons, rather than the CMB, and thus occurs again largely in the Klein-Nishina limit, at least until the particles reach an energy of $E_e \sim 10$ TeV or so. By this point, the cascade has traveled a distance of roughly ~ 100 Mpc, most of which was crossed since the last pair production event or two (see Figure 16.2). As the cascade leaves the Klein-Nishina regime for the second time, the resulting inverse Compton photons now carry less energy than their parent electrons by a factor of roughly $\langle E_\gamma / E_e \rangle \sim (E_e/10\,{\rm TeV})$. Over the next two or three cycles of pair production and inverse Compton scattering, the energies of the particles drop by roughly an order of magnitude as they propagate a distance of roughly 1 Gpc. It is at this point in the evolution that this electromagnetic cascade of TeV-scale photons reaches us.

Problem 16.4

Retaining the possibility of a finite value of $\sin^2\theta_{13}$ (while still keeping $\delta_{\rm CP} = 0$), the oscillation probability is given by:

$$P_{\nu_\alpha \to \nu_\beta} = \delta_{\alpha\beta} - 2 \sum_{i>j} U_{\alpha i} U_{\beta i} U_{\alpha j} U_{\beta j}.$$

For the $\nu_e \to \nu_e$ case, this becomes

$$P_{\nu_e \to \nu_e} = 1 - 2 \sum_{i>j} U_{ei}^2 U_{ej}^2 = 1 - 2U_{e3}^2 U_{e2}^2 - 2U_{e3}^2 U_{e1}^2 - 2U_{e2}^2 U_{e1}^2.$$

Using the entries from the PMNS matrix as given in Eq. 11.15 (for $\delta_{\rm CP} = 0$), this becomes

$$P_{\nu_e \to \nu_e} = 1 - 2s_{13}^2 s_{12}^2 c_{13}^2 - 2s_{13}^2 c_{12}^2 c_{13}^2 - 2s_{12}^2 c_{13}^4 c_{12}^2,$$

which for $\sin^2\theta_{13} = 0$ reduces to the expression given in Eq. 16.8. For $\sin^2\theta_{13} = 0.022$ and $\sin^2\theta_{12} = 0.31$, we obtain $P_{\nu_e \to \nu_e} \approx 0.54$. The other oscillation probabilities can be found following a similar procedure.

Problem 16.5

In this case, it simply follows that:

$$\phi_{\bar{\nu}_e} : \phi_{\bar{\nu}_\mu} : \phi_{\bar{\nu}_\tau} = P_{\bar{\nu}_e \to \bar{\nu}_e} : P_{\bar{\nu}_e \to \bar{\nu}_\mu} : P_{\bar{\nu}_e \to \bar{\nu}_\tau} \approx 0.57 : 0.20 : 0.23.$$

Problem 16.8

Numerically integrating

$$\xi \equiv \frac{\int_0^\infty f(z)\, H(z)^{-1}(1+z)^{-2}\, dz}{\int_0^\infty H(z)^{-1}(1+z)^{-1}\, dz}$$

for the case of $f(z) = (1+7.6z)/[1+(z/3.3)^{5/3}]$, and evaluating $H(z)$ using $\Omega_M = 0.315$ and $\Omega_\Lambda = 0.685$, we find $\xi \approx 2.8$.

Chapter 17

Problem 17.3

To find the eigenstates, we solve $|\mathcal{M}^\dagger_{\chi^\pm}\mathcal{M}_{\chi^\pm} - I\, m^2_{\chi^\pm}| = 0$ for $m^2_{\chi^\pm}$:

$$\begin{aligned}
0 &= \left| \begin{pmatrix} M_2 & \sqrt{2}\, m_W \cos\beta \\ \sqrt{2}\, m_W \sin\beta & \mu \end{pmatrix} \begin{pmatrix} M_2 & \sqrt{2}\, m_W \sin\beta \\ \sqrt{2}\, m_W \cos\beta & \mu \end{pmatrix} - \begin{pmatrix} m^2_{\chi^\pm} & 0 \\ 0 & m^2_{\chi^\pm} \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} M_2^2 + 2m_W^2 \cos^2\beta - m^2_{\chi^\pm} & \sqrt{2}\, m_W M_2 \sin\beta + \sqrt{2} m_W \mu \cos\beta \\ \sqrt{2}\, M_2 m_W \sin\beta + \sqrt{2} m_W \mu \cos\beta & \mu^2 + 2m_W^2 \sin^2\beta - m^2_{\chi^\pm} \end{pmatrix} \right| \\
&= (M_2^2 + 2m_W^2\cos^2\beta - m^2_{\chi^\pm})(\mu^2 + 2m_W^2 \sin^2\beta - m^2_{\chi^\pm}) \\
&\quad - (\sqrt{2}\, m_W M_2 \sin\beta + \sqrt{2} m_W \mu \cos\beta)(\sqrt{2}\, M_2 m_W \sin\beta + \sqrt{2} m_W \mu \cos\beta). \\
&= m^4_{\chi^\pm} + m^2_{\chi^\pm}(-M_2^2 - \mu^2 - 2m_W^2) + (M_2^2 + 2m_W^2\cos^2\beta)(\mu^2 + 2m_W^2\sin^2\beta) \\
&\quad - 2m_W^2(M_2\sin\beta + \mu\sin\beta)^2 \\
&= m^4_{\chi^\pm} + m^2_{\chi^\pm}(-M_2^2 - \mu^2 - 2m_W^2) + (\mu M_2 - 2m_W^2 \sin\beta\cos\beta)^2 \\
&= m^4_{\chi^\pm} + m^2_{\chi^\pm}(-M_2^2 - \mu^2 - 2m_W^2) + (\mu M_2 - m_W^2 \sin 2\beta)^2.
\end{aligned}$$

Applying the quadratic formula, we obtain

$$m^2_{\chi^\pm} = \frac{1}{2}\left[M_2^2 + \mu^2 + 2m_W^2 \mp \sqrt{(M_2^2 + \mu^2 + 2m_W^2)^2 - 4(\mu M_2 - m_W^2 \sin 2\beta)^2} \right].$$

Next, to find the chargino masses in the $|M_2 \pm \mu| \gg m_W$ limit, we focus on the piece of Eq. 17.16 in the square root:

$$\begin{aligned}
&(M_2^2 + \mu^2 + 2m_W^2)^2 - 4(\mu M_2 - m_W^2 \sin 2\beta)^2 \\
&= (M_2^4 + \mu^4 + 2M_2^2\mu^2)\left(1 + \frac{2m_W^2}{(M_2^2+\mu^2)}\right)^2 - 4\mu^2 M_2^2 \left(1 - \frac{m_W^2 \sin 2\beta}{\mu M_2}\right)^2
\end{aligned}$$

$$
\begin{aligned}
&\approx (M_2^4+\mu^4+2M_2^2\mu^2)\left(1+\frac{4m_W^2}{(M_2^2+\mu^2)}\right)-4\mu^2M_2^2\left(1-\frac{2m_W^2\sin 2\beta}{\mu M_2}\right)\\
&=M_2^4+\mu^4+2M_2^2\mu^2+(M_2^4+\mu^4+2M_2^2\mu^2)\left(\frac{4m_W^2}{(M_2^2+\mu^2)}\right)\\
&\quad-4\mu^2M_2^2+4\mu^2M_2^2\left(\frac{2m_W^2\sin 2\beta}{\mu M_2}\right)\\
&=M_2^4+\mu^4-2M_2^2\mu^2+(M_2^4+\mu^4+2M_2^2\mu^2)\left(\frac{4m_W^2}{(M_2^2+\mu^2)}\right)+8\mu M_2m_W^2\sin 2\beta\\
&=(M_2^2-\mu^2)^2+(M_2^2+\mu^2)4m_W^2+8\mu M_2m_W^2\sin 2\beta\\
&=(M_2^2-\mu^2)^2\left[1+\frac{(M_2^2+\mu^2)4m_W^2+8\mu M_2m_W^2\sin 2\beta}{(M_2^2-\mu^2)^2}\right],
\end{aligned}
$$

where in the third line I dropped the terms proportional to m_W^4. Inserting this into the square root, we obtain the following expression:

$$
\begin{aligned}
2m_{\chi^\pm}^2&=M_2^2+\mu^2+2m_W^2\mp|M_2^2-\mu^2|\left[1+\frac{(M_2^2+\mu^2)4m_W^2+8\mu M_2m_W^2\sin 2\beta}{(M_2^2-\mu^2)^2}\right]^{1/2}\\
&\approx M_2^2+\mu^2+2m_W^2\mp|M_2^2-\mu^2|\left[1+\frac{(M_2^2+\mu^2)2m_W^2+4\mu M_2m_W^2\sin 2\beta}{(M_2^2-\mu^2)^2}\right]\\
&=M_2^2+\mu^2+2m_W^2\mp\left[|M_2^2-\mu^2|+\frac{(M_2^2+\mu^2)2m_W^2+4\mu M_2m_W^2\sin 2\beta}{|M_2^2-\mu^2|}\right],
\end{aligned}
$$

where in the second line I performed a Taylor series expansion. Taking first the case of the $+$ sign and $M_2^2-\mu^2>0$ (or the $-$ sign and $M_2^2-\mu^2<0$), this expression reduces to

$$
\begin{aligned}
2m_{\chi^\pm}^2&=M_2^2+\mu^2+2m_W^2+M_2^2-\mu^2+\frac{(M_2^2+\mu^2)2m_W^2+4\mu M_2m_W^2\sin 2\beta}{(M_2^2-\mu^2)}\\
&=2M_2^2+2m_W^2+\frac{(M_2^2+\mu^2)2m_W^2+4\mu M_2m_W^2\sin 2\beta}{(M_2^2-\mu^2)}.
\end{aligned}
$$

Taking the square root, we arrive at

$$
\begin{aligned}
m_{\chi^\pm}&=\left[M_2^2+m_W^2\left(1+\frac{M_2^2+\mu^2+2\mu M_2\sin 2\beta}{(M_2^2-\mu^2)}\right)\right]^{1/2}\\
&=M_2\left[1+\frac{m_W^2}{M_2^2}\left(\frac{M_2^2-\mu^2+M_2^2+\mu^2+2\mu M_2\sin 2\beta}{(M_2^2-\mu^2)}\right)\right]^{1/2}\\
&=M_2\left[1+\frac{2m_W^2}{M_2}\left(\frac{M_2+\mu\sin 2\beta}{M_2^2-\mu^2}\right)\right]^{1/2}
\end{aligned}
$$

$$\approx M_2\left[1+\frac{m_W^2}{M_2}\left(\frac{M_2+\mu\sin 2\beta}{M_2^2-\mu^2}\right)\right]$$

$$=M_2-\frac{m_W^2(M_2+\mu\sin 2\beta)}{\mu^2-M_2^2},$$

where in the fourth line I performed a Taylor series expansion. Repeating the later steps of this calculation for the case of the $-$ sign and $M_2^2-\mu^2>0$ (or the $+$ sign and $M_2^2-\mu^2<0$), the expression instead reduces to

$$2m_{\chi^\pm}^2=M_2^2+\mu^2+2m_W^2-M_2^2+\mu^2-\frac{(M_2^2+\mu^2)2m_W^2+4\mu M_2 m_W^2\sin 2\beta}{(M_2^2-\mu^2)}$$

$$=2\mu^2+2m_W^2-\frac{(M_2^2+\mu^2)2m_W^2+4\mu M_2 m_W^2\sin 2\beta}{(M_2^2-\mu^2)}.$$

After the square root, we obtain

$$m_{\chi^\pm}=\left[\mu^2+m_W^2\left(1-\frac{M_2^2+\mu^2+2\mu M_2\sin 2\beta}{(M_2^2-\mu^2)}\right)\right]^{1/2}$$

$$=|\mu|\left[1+\frac{m_W^2}{\mu^2}\left(\frac{M_2^2-\mu^2-M_2^2-\mu^2-2\mu M_2\sin 2\beta}{(M_2^2-\mu^2)}\right)\right]^{1/2}$$

$$=|\mu|\left[1-\frac{2m_W^2}{\mu}\left(\frac{\mu+M_2\sin 2\beta}{M_2^2-\mu^2}\right)\right]^{1/2}$$

$$\approx|\mu|\left[1-\frac{m_W^2}{\mu}\left(\frac{\mu+M_2\sin 2\beta}{M_2^2-\mu^2}\right)\right]$$

$$=|\mu|+\frac{\mu}{|\mu|}\frac{m_W^2(\mu+M_2\sin 2\beta)}{\mu^2-M_2^2}.$$

Problem 17.4

To find the eigenstates, we solve $|\mathcal{M}_\nu - I\,m_\nu|=0$ for m_ν:

$$0=\left|\begin{pmatrix}-m_\nu & \mu\\ \mu & M-m_\nu\end{pmatrix}\right|$$

$$=m_\nu M+m_\nu^2-\mu^2.$$

Applying the quadratic formula,

$$m_\nu=\frac{M\mp\sqrt{M^2+4\mu^2}}{2}.$$

This can be written as

$$\begin{aligned} m_\nu &= \frac{M}{2} \mp \frac{M}{2}\left(1+\frac{4\mu^2}{M^2}\right)^{1/2} \\ &\approx \frac{M}{2} \mp \frac{M}{2}\left(1+\frac{2\mu^2}{M^2}\right) \\ &= \frac{M}{2} \mp \left[\frac{M}{2}+\left(\frac{\mu^2}{M}\right)\right] \\ &= -\left(\frac{\mu^2}{M}\right) \quad \text{or} \quad M+\left(\frac{\mu^2}{M}\right) \approx M, \end{aligned}$$

where in the second line I performed a Taylor series expansion in the $\mu \ll M$ limit.

Problem 17.5

The left-hand side of Eq. 17.75 can be written as

$$\begin{aligned} \frac{dr}{d\ln a} &= n_s \frac{d(1/n_a)}{d\ln a} + \frac{1}{n_a}\frac{dn_s}{d\ln a} = n_s a \frac{d(1/n_a)}{da} + \frac{a}{n_a}\frac{dn_s}{da} \\ &= n_s a \frac{d}{da}\left(\frac{2\pi^2}{3\zeta(3)T^3}\right) + \frac{1}{n_a H}\frac{dn_s}{dt} = n_s a\left(\frac{-2\pi^2}{\zeta(3)T^4}\frac{dT}{da}\right) + \frac{1}{n_a H}\frac{dn_s}{dt} \\ &= n_s a\left(\frac{-2\pi^2}{\zeta(3)T^4}\left[-\frac{T}{a}-\frac{T}{3g_{\star,S}}\frac{dg_{\star,S}}{da}\right]\right) + \frac{1}{n_a H}\frac{dn_s}{dt} \\ &= \frac{n_s}{n_a}\left[\frac{a}{g_{\star,S}}\frac{dg_{\star,S}}{da}+3\right] + \frac{1}{n_a H}\frac{dn_s}{dt}, \end{aligned}$$

while the right-hand side can be written as

$$\frac{\gamma}{H} + r\frac{d\ln g_{\star,S}}{d\ln a} = \frac{\sin^2 2\theta_{\rm eff}\Gamma}{2H} + \frac{n_s}{n_a}\frac{d\ln g_{\star,S}}{d\ln a}.$$

Combing these and multiplying both sides by $n_a H$ yields the following:

$$n_s H\left[\frac{a}{g_{\star,S}}\frac{dg_{\star,S}}{da}+3\right] + \frac{dn_s}{dt} = n_a\frac{\sin^2 2\theta_{\rm eff}\Gamma}{2} + n_s H\frac{d\ln g_{\star,S}}{d\ln a},$$

which reduces to

$$\frac{dn_s}{dt} = -3n_s H + n_a\frac{\sin^2 2\theta_{\rm eff}\Gamma}{2}.$$

This is equivalent to the Boltzmann equation as expressed in Eq. 17.29.

Problem 17.7

We start by taking the derivative of Eq. 17.42:

$$\begin{aligned}\frac{\partial V}{\partial \phi} &\approx \frac{-m_\pi^2 f_\pi^2}{2}\left[1-\frac{4m_u m_d}{(m_u+m_d)^2}\sin^2(N\phi/2f_{\rm PQ})\right]^{-1/2} \\ &\quad\times\left[-2\frac{4m_u m_d}{(m_u+m_d)^2}\sin(N\phi/2f_{\rm PQ})\,\cos(N\phi/2f_{\rm PQ})\frac{N}{2f_{\rm PQ}}\right] \\ &= \frac{2m_\pi^2 f_\pi^2\, m_u m_d N}{(m_u+m_d)^2 f_{\rm PQ}}\left[1-\frac{4m_u m_d}{(m_u+m_d)^2}\sin^2(N\phi/2f_{\rm PQ})\right]^{-1/2} \\ &\quad\times\left[\sin(N\phi/2f_{\rm PQ})\,\cos(N\phi/2f_{\rm PQ})\right].\end{aligned}$$

Taking the second derivative, and dropping all of the terms proportional to a sine function (which are zero when evaluated at $\phi = 0$), we arrive at

$$\begin{aligned}\frac{\partial^2 V}{\partial \phi^2}(\phi=0) &\approx \frac{2m_\pi^2 f_\pi^2\, m_u m_d N}{(m_u+m_d)^2 f_{\rm PQ}}\left[0+\left(1-0\right)^{-1/2}\left(\cos^2(N\phi/2f_{\rm PQ})\frac{N}{2f_{\rm PQ}}-0\right)\right] \\ &= \frac{2m_\pi^2 f_\pi^2\, m_u m_d N}{(m_u+m_d)^2 f_{\rm PQ}}\left[\frac{N}{2f_{\rm PQ}}\right] \\ &= \frac{m_\pi^2 f_\pi^2\, m_u m_d}{(m_u+m_d)^2\,(f_{\rm PQ}/N)^2},\end{aligned}$$

which matches the result reported in Eq. 17.43.

Problem 17.8

For such a large value of $f_{\rm PQ}$, the axions decouple prior to $T \sim \Lambda_{\rm QCD}$, and thus $g_{\star,S}(T_{\rm dc}) \sim 10^2$. It thus follows from Eq. 17.54 that

$$\begin{aligned}\Delta N_{\rm eff} &\approx 0.15\left(\frac{30}{100}\right)^{4/3}\times\left(\frac{100}{g_{\star,S}(T_{\rm dm})}\right)^{4/3} \\ &\approx 0.03\times\left(\frac{100}{g_{\star,S}(T_{\rm dm})}\right)^{4/3}.\end{aligned}$$

Problem 17.9

Near the minimum at $\phi = 0$, we can apply the small angle approximation, $\sin(N\phi/2f_{\rm PQ}) \approx N\phi/2f_{\rm PQ}$, to write Eq. 17.42 as follows:

$$V(\phi) \approx m_\pi^2 f_\pi^2 - m_\pi^2 f_\pi^2\left[1-\frac{4m_u m_d}{(m_u+m_d)^2}(N\phi/2f_{\rm PQ})^2\right]^{1/2}.$$

Applying a Taylor series expansion, this becomes

$$V(\phi) \approx m_\pi^2 f_\pi^2 - m_\pi^2 f_\pi^2 \left[1 - \frac{2m_u m_d}{(m_u + m_d)^2} (N\phi/2f_{\rm PQ})^2\right]$$

$$= m_\pi^2 f_\pi^2 \left[\frac{2m_u m_d}{(m_u + m_d)^2} (N\phi/2f_{\rm PQ})^2\right].$$

Using the expression for m_a given in Eq. 17.43, this becomes

$$V(\phi) \approx \frac{m_a^2 \phi^2}{2},$$

which is equal to $V(\phi) \approx (m_a^2/2)(f_{\rm PQ}/N)^2\, \theta_{\rm QCD}^2$.

Problem 17.10

The mass contained within the horizon at $T = f_{\rm PQ} = 10^{12}$ GeV is given by

$$M_{\rm hor} \sim \frac{M_{\rm Pl}^2}{H} = \frac{M_{\rm Pl}^2}{(8\pi^3 g_\star f_{\rm PQ}^4/90 M_{\rm Pl}^2)^{1/2}} \approx 10^{32}\,{\rm GeV} \sim 2 \times 10^5\,{\rm kg}.$$

Thus even the smallest galaxies contain matter from numerous regions that were causally disconnected at the time of Peccei-Quinn symmetry breaking.

Chapter 18

Problem 18.1

Expanding the sin and cos in Taylor series, the uniform sphere form factor in the $qr_{\rm max} \ll 1$ limit is given by

$$F_{\rm SI}(q) \approx A\, \frac{3[qr_{\rm max} - (qr_{\rm max})^3/3!] - 3qr_{\rm max}[1 - (qr_{\rm max})^2/2!]}{(qr_{\rm max})^3}$$

$$= A\, \frac{-3(qr_{\rm max})^3/6 + 3(qr_{\rm max})^3/2}{(qr_{\rm max})^3} = A.$$

Problem 18.2

If xenon nuclei were 10 times larger in radius, the spin-independent form factor would suppress the scattering rate down to recoil momenta that are 10 times smaller than usual. So instead of suppressing scattering with recoil energies greater than $10\,{\rm keV} \times (100/132)^{1/3}$, scattering would be suppressed if

$$E_R = \frac{\vec{q}^{\,2}}{2m_N} \gtrsim 0.1^2 \times 10\,{\rm keV} \times \left(\frac{100}{132}\right)^{1/3} \approx 0.09\,{\rm keV}.$$

Given that this is well below the threshold for detection, we conclude that observable events would not be able to benefit from coherent scattering, and thus the relevant event rates would not be enhanced by the factor of A^2 that is ordinarily found in the case of spin-independent scattering.

Problem 18.3

Since third generation quarks are not present in nuclei, this dark matter candidate can only scatter with gluons, through a bottom or top quark loop.

From Eq. 18.21, we know that to obtain the measured dark matter abundance, we must require $g_X g_q / m_M^2 \approx 4.4 \times 10^{-5}\,\text{GeV}^{-1}/(m_X\, n_q^{1/2})$, where n_q is either 1 or 2, depending on whether the dark matter is heavy enough to annihilate to a pair of top quarks. To calculate the elastic scattering cross section of this dark matter candidate, we apply Eqs. 18.17 and 18.18, replacing the mediator's coupling to quarks, $m_q \lambda_q$, with g_q. This leads to $f_{p,n} = (2/27) f_{TG}(\lambda_b + \lambda_t) = (2/27) f_{TG}[(g_b/m_b) + (g_t/m_t)] \approx (2/27) f_{TG}\, g_b/m_b$. This, in turn, leads to the following spin-independent scattering cross section:

$$\begin{aligned}
\sigma_{p,n} &\approx \frac{\mu_{Xp,n}^2 g_X^2}{\pi m_M^4} \frac{2 g_b}{27 m_b} f_{TG} \\
&\approx \frac{\mu_{Xp,n}^2}{\pi} \frac{2}{27 m_b} f_{TG} \left(\frac{4.4 \times 10^{-5}\,\text{GeV}^{-1}}{m_X n_q^{1/2}} \right)^2 \\
&\approx 1.7 \times 10^{-45}\,\text{cm}^2 \times \left(\frac{1\,\text{TeV}}{m_X} \right)^2 \left(\frac{2}{n_q} \right).
\end{aligned}$$

Comparing this to the constraints presented in Figure 18.2, this dark matter candidate is viable only if it is heavier than $m_X \sim 1.9\,\text{TeV}$.

Problem 18.6

The minimum velocity required to exceed the threshold for this process is found in the case in which the momenta of the incoming particle, outgoing particle, and target nuclei are all aligned in the same direction. In this case, and including the difference in mass between the incoming and outgoing particles, we can rewrite Eq. 18.2 as follows:

$$\begin{aligned}
\frac{\vec{p}^{\,2}}{2m_\chi} &= \frac{(\vec{p} - \vec{q})^2}{2(m_\chi + \delta)} + \frac{\vec{q}^{\,2}}{2m_N} + \delta \\
&= \frac{p^2 + q^2 - 2pq}{2(m_\chi + \delta)} + \frac{q^2}{2m_N} + \delta,
\end{aligned}$$

where p and q are the magnitudes of the momenta of the incoming particle and the recoiling target, and $\delta = 100\,\text{keV}$ is the mass splitting between the incoming and outgoing particles. Solving this for q using the quadratic equation, we find that the result is real only if

$$p^2 m_\chi^2 m_N^2 - m_\chi m_N (m_N + m_\chi + \delta)(-\delta p^2 + 2\delta m_\chi (m_\chi + \delta)) > 0,$$

which we can solve for p to obtain the following condition:

$$\begin{aligned}
p^2 = m_\chi^2 v^2 &> \frac{m_\chi m_N (m_N + m_\chi + \delta)(2\delta m_\chi (m_\chi + \delta))}{m_\chi^2 m_N^2 + m_\chi m_N \delta (m_N + m_\chi + \delta)} \\
&> \frac{(m_N + m_\chi + \delta)(2\delta m_\chi (m_\chi + \delta))}{m_\chi m_N + \delta (m_N + m_\chi + \delta)}
\end{aligned}$$

$$\longrightarrow \quad v^2 \gtrsim \frac{(m_N + m_\chi)2\delta}{m_\chi m_N} = \frac{2\delta}{\mu_{\chi N}},$$

where in the last line I have dropped higher order terms in δ. For $m_N = 132\, m_{p,n}$, $m_\chi = 1$ TeV, and $\delta = 100$ keV, this process requires the dark matter particle to be traveling at a velocity that is no less than $v \gtrsim 0.0013c \approx 400$ km/s.

Problem 18.7

To cover the mass range between 10^{-6} and 10^{-5} eV in steps of $\Delta\nu/m_a = 10^{-5}$ (and thus $\Delta m_a/m_a = 2\pi \times 10^{-5}$) would require $\sim$36,000 steps. From Eq. 18.29, it will take $t \sim 1{,}800\,\text{s} \times (10^{-5}\,\text{eV}/m_a)^5$ to achieve $S/N = 2$ at each step, for a total time of $\sim 6 \times 10^{11}$ s $\sim$ 20,000 yr. Note that the time required is dominated by the lightest axions, since $S/N \propto m_a^{5/2}$.

Chapter 19

Problem 19.1

Between the time of kinetic decoupling and recombination, the temperature of the dark matter decreases with a^2, such that $T_{\rm DM} \approx T_{\rm kd} \times (T_{\rm dc}/T_{\rm kd})^2 \approx (0.25\,\text{eV})^2/(10\,\text{keV}) \approx 7 \times 10^{-6}$ eV. This corresponds to a typical velocity of $v^2 \sim T_{\rm DM}/m_X$, and thus to $\langle\sigma v\rangle \sim 2.2 \times 10^{-26}\,\text{cm}^3/\text{s} \times (7 \times 10^{-6}\,\text{eV}/m_X/0.1)$. Applying the constraint given in Eq. 19.3, we have

$$\frac{2.2 \times 10^{-26}\,\text{cm}^3/\text{s} \times 7 \times 10^{-5}\,\text{eV}}{m_X}\,\frac{0.5}{m_X} \lesssim 3.2 \times 10^{-37}\,\text{cm}^3/\text{s/eV},$$

which reduces to $m_X \gtrsim 1.5$ keV. The lesson here is that the CMB does not meaningfully constrain a purely p-wave annihilating thermal relic.

Problem 19.2

A line of sight directed along an angle θ away from the center of the sphere passes through the surface of the sphere at $l = d\cos\theta + \sqrt{r^2 - d^2\sin^2\theta}$ and again at $l = d\cos\theta - \sqrt{r^2 - d^2\sin^2\theta}$, for a total distance within the sphere of $x(\theta) = 2\sqrt{r^2 - d^2\sin^2\theta}$. The integral over the line of sight and solid angle can then be evaluated as follows:

$$\begin{aligned} J &\equiv \int_{\Delta\Omega}\int_{los} \rho_X^2(l,\Omega)\, dl\, d\Omega = \rho^2 \int_0^{\arcsin(r/d)} \int_0^{2\pi} 2\sqrt{r^2 - d^2\sin^2\theta}\, d\phi\, \sin\theta\, d\theta \\ &= 4\pi\rho^2 \int_0^{\arcsin(r/d)} \sqrt{r^2 - d^2\sin^2\theta}\, \sin\theta\, d\theta = \frac{4\pi r^3 \rho^2}{3d^2}. \end{aligned}$$

Problem 19.7

In the limit of a uniform source term, the steady-state solution to the diffusion-loss equation is, in this case, given by

$$0 = \frac{d}{dE}\left[b(E)\,\frac{dN_e}{dE_e}\right] + Q(E_e),$$

where $Q(E_e) = (1/2)\,\langle\sigma\,v\rangle\,(\rho_X/m_X)^2\,\delta(E_e - m_X)$, and $b \equiv -dE_e/dt \approx 10^{-16}\,\mathrm{GeV/s} \times (E_e/\mathrm{GeV})^2\,(\rho_{\rm rad} + \rho_B)/(\mathrm{eV\,cm^{-3}})$.

For $E_e < m_X$, the last term vanishes, and the solution is given by $dN_e/dE_e \propto E_e^{-2}$. To normalize this spectrum, we need to evaluate the flux at $E_e = m_X$, where the solution is $dN_e/dE_e = Q(E_e)/b(E_e)$. Combining this information, the overall solution is

$$\frac{dN_e}{dE_e} \approx \frac{\langle\sigma\,v\rangle}{2}\,\frac{\rho_X^2}{m_X^2}\,\frac{1}{10^{-10}\,\mathrm{GeV/s}}\left(\frac{\mathrm{eV/cm^3}}{\rho_{\rm rad}+\rho_B}\right)\left(\frac{m_X}{E_e}\right)^2$$

$$\approx 1.8\times 10^{-23}\,\mathrm{GeV^{-1}\,cm^{-3}} \times \left(\frac{\mathrm{eV/cm^3}}{\rho_{\rm rad}+\rho_B}\right)\left(\frac{\mathrm{TeV}}{E_e}\right)^2,$$

up to $E_e = m_X$.

To convert this number density per energy into a flux, we multiply by $c/4\pi$, for a final result of $\sim 4\times 10^{-14}\,\mathrm{GeV^{-1}\,cm^{-2}\,s^{-1}\,sr^{-1}} \times (\mathrm{TeV}/E_e)^2\,(\mathrm{eV\,cm^{-3}})/(\rho_{\rm rad}+\rho_B)$.

Chapter 20

Problem 20.1

For a fixed target:

$$\begin{aligned}
E_{\rm cm} &= \sqrt{E_{\rm tot}^2 - p_{\rm tot}^2} = \sqrt{(\gamma_{\rm beam} m_{\rm beam} + m_{\rm target})^2 - (\gamma_{\rm beam} m_{\rm beam}\, v_{\rm beam})^2} \\
&= \sqrt{\gamma_{\rm beam}^2 m_{\rm beam}^2 + m_{\rm target}^2 + 2\gamma_{\rm beam} m_{\rm beam} m_{\rm target} - \gamma_{\rm beam}^2 m_{\rm beam}^2 v_{\rm beam}^2} \\
&= \sqrt{\gamma_{\rm beam}^2 m_{\rm beam}^2 (1 - v_{\rm beam}^2) + m_{\rm target}^2 + 2\gamma_{\rm beam} m_{\rm beam} m_{\rm target}} \\
&= \sqrt{\gamma_{\rm beam}^2 m_{\rm beam}^2 \gamma_{\rm beam}^{-2} + m_{\rm target}^2 + 2\gamma_{\rm beam} m_{\rm beam} m_{\rm target}} \\
&= \sqrt{m_{\rm beam}^2 + m_{\rm target}^2 + 2E_{\rm beam} m_{\rm target}}.
\end{aligned}$$

And for colliding beams:

$$\begin{aligned}
E_{\rm cm} &= \sqrt{E_{\rm tot}^2 - p_{\rm tot}^2} = \sqrt{(\gamma_1 m_1 + \gamma_2 m_2)^2 - (\gamma_1 m_1 v_1 - \gamma_2 m_2 v_2)^2} \\
&= \sqrt{(\gamma_1^2 m_1^2 + \gamma_2^2 m_2^2 + 2\gamma_1\gamma_2 m_1 m_2) - (\gamma_1^2 m_1^2 v_1^2 + \gamma_2^2 m_2^2 v_2^2 - 2\gamma_1\gamma_2 m_1 m_2 v_1 v_2)} \\
&= \sqrt{\gamma_1^2 m_1^2 (1 - v_1^2) + \gamma_2^2 m_2^2 (1 - v_2^2) + 2\gamma_1\gamma_2 m_1 m_2 + 2\gamma_1\gamma_2 m_1 m_2 v_1 v_2} \\
&= \sqrt{\gamma_1^2 m_1^2 \gamma_1^{-2} + \gamma_2^2 m_2^2 \gamma_2^{-2} + 2\gamma_1\gamma_2 m_1 m_2 + 2\gamma_1\gamma_2 m_1 m_2 v_1 v_2} \\
&= \sqrt{m_1^2 + m_2^2 + 2E_1 E_2 + 2p_1 p_2}.
\end{aligned}$$

Problem 20.3

The invariant mass of the electron-positron pair is given by:

$$\begin{aligned} m &= [(E_1 + E_2)^2 - (\vec{p}_1 + \vec{p}_2)^2]^{1/2} \\ &\approx [(|\vec{p}_1| + |\vec{p}_2|)^2 - (\vec{p}_1 + \vec{p}_2)^2]^{1/2} \\ &\approx [(100 + 51.6\,\text{GeV})^2 - (110\,\text{GeV})^2 - (50.6\,\text{GeV})^2]^{1/2} \\ &\approx 91.2\,\text{GeV}. \end{aligned}$$

Since the invariant mass of these particles is close to the value of m_Z, it is likely that they were produced through the decay of a Z boson.

Problem 20.4

With a lifetime of $\tau \approx 1.3 \times 10^{-12}$ s, the typical 100 GeV b quark travels roughly $d \sim (100/4.2) \times c\tau \sim 0.9$ cm before decaying. The finite lifetime of this particle thus causes its decay to occur at a location that is nonnegligibly displaced from the accelerator's collision point. It is often possible to use this feature to distinguish a b-jet from other kinds of jets at the LHC. This process is known as "b-tagging."

Problem 20.5

Following Sec. 3.5, the vertex factors that go into the amplitude for this process are $(g_s/2)^2 (y_t)^2 (Q_t g_2 \sin\theta_W)^2$, where $y_t \sim 1$ is the coupling of the Higgs boson to a pair of top quarks and $Q_t = 2/3$. The propagators lead to two factors of $1/(p^2 - m_t^2)$ and one of $1/(p^2 - m_h^2)$, which are collectively on the order of E_{tot}^{-6}. Since the amplitude should be unitless, we multiply the overall result by a factor of E_{tot}^{-6}, leaving us with $\mathcal{M} \sim g_s^2 y_t^2 g_2^2 \sin^2\theta_W/9$. Taking into account that the top quark loop connected to the outgoing photons could be of any color or anticolor, we include an additional factor of 6, and thus arrive at an estimate of $|\mathcal{M}|^2 \sim 4 g_s^4 y_t^4 g_2^4 \sin^4\theta_W/9$.

Inserting this into Eq. 3.15, we arrive at the following estimate for the cross section:

$$\begin{aligned} \sigma_{gg\to h\to\gamma\gamma} &\sim 4\pi\, \frac{(1/2)|\mathcal{M}|^2}{64\pi^2\, E_{\text{tot}}^2}, \\ &\sim \frac{g_s^4 y_t^4 g_2^4 \sin^4\theta_W}{72\pi\, E_{\text{tot}}^2} \sim 0.36\,\text{pb} \times \left(\frac{300\,\text{GeV}}{E_{\text{tot}}}\right)^2. \end{aligned}$$

This is of the same order of magnitude as $\sigma_{pp\to h} \times BF(h \to \gamma\gamma) \approx 60\,\text{pb} \times 0.002 \approx 0.12\,\text{pb}$.

Problem 20.6

For the LHC's design luminosity of $10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, a rate of 10^4 events per year corresponds to a cross section of $\sigma \sim 3 \times 10^{-2}$ pb. For gluino pair production, for example, this cross section corresponds to roughly $m_{\tilde{g}} \sim 1{,}400$ GeV. For charginos, it corresponds to $m_{\tilde{\chi}} \sim 500$ GeV.

Index

Page numbers in italics refer to figures or tables.

acoustic oscillations, 93–94, 95; baryon abundance and, 100; CMB polarization and, 102; dark energy and, 235, *235*, 236; dark matter and, 100–101; large-scale structure and, 153, 155, 168–70, *169*; stable under gravity, 153
acoustic peaks, 95–96, 98, 100; accelerating expansion and, 231; CMB polarization and, 102; suppressed by dark matter, 101; transfer function and, 168–69, *169*
active galactic nuclei (AGN), 260–64; cosmic-ray acceleration and, 247; diffuse high-energy neutrino flux from, *260*, 263, 264; jets of, 261–64, *262*; neutrinos from non-blazar, *260*, 263, 264; ultra-high energy cosmic rays and, 249, *249*, 261. *See also* blazars
adiabatic density perturbations, 170–71
A-funnel region, 281
Alpher, Ralph, 108
amplitudes, 33, 39–40
angular diameter distance, 26, *27*, 28; dark energy equation of state and, 234, 235, *235*
angular power spectrum of CMB: adiabatic initial conditions and, 171; *B*-mode polarization in, 101–2, 220; cosmic strings and, 206; curvature of universe and, 98, *99*, 100; dark matter energy density and, 234; diffusion damping of, 97–98; Fourier modes and, 92; integrated Sachs-Wolfe (ISW) effect, 95; measured values of, 92, *93*; neutrino masses and, 97–98, 104; predicted values of, 88–92; Sachs-Wolfe plateau, 94–95; secondary effects on, 98; slow-roll parameters and, 219–20; statistical properties of, 157–58; suppressed in reionization epoch, 83–84; time of last scattering and, 98. *See also* acoustic peaks
anisotropies of CMB, 88, 89, *89*, 94, 101; inflation and, 226; isocurvature perturbations and, 297; secondary, 234n; suppressed in reionization epoch, 83–84
annihilation *J*-factors, 317, *318*, 318–19, 320
anomalies, in quantum field theory, 289
anthropic principle: axion misalignment angle and, 297; dark energy and, 236–37
anti-deuterons, 323
anti-helium nuclei, 323
antimatter: arguments for lack of, 183; from dark matter annihilation or decay, 321. *See also* baryon asymmetry; matter-antimatter asymmetry
antinuclei in cosmic rays, 321, 322–23
antiprotons in cosmic rays, *320*, 321, 322–23
antiquarks, in proton colliders, 328
asymptotic freedom, 37n
atoms, formation of, 6, 7, 54, 79. *See also* recombination
autocorrelation function, 159
axinos, 274
axion color anomaly, 291
axion decay constant, 289
axion electromagnetic anomaly, 291
axion-like particles (ALPs), 291–92
axion production in early universe, 292–300, *294*; energy density of, 293, 295–97, 298;

axion production in early universe (*continued*) number density of, 295, 296; thermal, 292–93, *294*; from topological defects, *294*, 298–300; from vacuum misalignment, 293–98, *294*, 299–300
axions, 288–300; astrophysical constraints on, 290; broken $U(1)_{PQ}$ symmetry and, 289, 291, 297, 298, 300; coupling to two photons, 290, 291, 311–12; as dark matter candidate, 291, 293, 297, 300; as dark radiation, 105; DFSZ model and, 291, 300; direct detection of, 311–12; KSVZ model and, 290–91, 300; lifetime of, 291; mass of, 289–90, 291, 293, 300; mass with temperature dependence, 294–96, 299; Peccei-Quinn mechanism and, 10, 288–92, 297–98, 300. *See also* axion production in early universe

barn, 330n
baryogenesis. *See* baryon asymmetry
baryon abundance: from CMB measurements, *99*, 100, 103; in MACHOs, 125; matter-antimatter asymmetry and, 139; primordial element abundances and, 119, *120*; from primordial helium mass fraction, 116; primordial lithium-7 and, 119, *120*
baryon acoustic oscillations. *See* acoustic oscillations
baryon asymmetry: by electroweak baryogenesis, 197–202; generated in first second, 184; in grand unified theories, 185–90; by leptogenesis, 194–97; Sakharov conditions and, 184–85; sphalerons and, 185, 190–94, 197. *See also* matter-antimatter asymmetry
baryonic Jeans length, 153
baryonic Jeans mass, 154
baryon-minus-lepton number ($B - L$): conserved in some GUTs, 188; sphalerons and, 190–91, *191*, 192, 194; vacuum transitions and, 190–91
baryon number: of asymmetric dark matter candidate, 140–41; *R*-parity and, 274. *See also* baryon-minus-lepton number ($B - L$); baryon-plus-lepton number ($B + L$)
baryon number conservation, 36; violated by first Sakharov condition, 184; violated in GUTs, 185, 186, 187
baryon-photon fluid: acoustic oscillations in, 93–94, 95, 102, 104, 153, 168–70, *169*; CMB polarization and, 102; dark matter and, 100–101; density perturbations in, 156, 157
baryon-plus-lepton number ($B + L$): sphalerons and, 190, 191, 192, 194
baryons, 37
Bekenstein, Jacob, 127
beryllium, *117*, 117–18, 119
Bethe, Hans, 108n
Big Bang nucleosynthesis (BBN), 5, 7, 108–23; beyond helium, 116–18, *117*; Boltzmann equation and, 61, 111; constraints on energy injection and, 121–22; dark matter freeze out and, 133; dark radiation and, 119, 121; equilibrium nuclear abundances and, 112–14, *114*; evolution of mass fractions during, *117*; of hydrogen and helium, 114–16, *115*; lack of observations prior to, 8, 332; limits on formation of heavy nuclei, 116–17; main processes in, 114–16, *115*; neutron-proton freeze out and, 109–12, *111*, 332; primordial element abundances, 118–19; unresolved lithium problem in, 119
Big Bang singularity, 4
Big Crunch, 24
binos, 274–75
black holes: active galactic nuclei and, 260, *262*; at center of Milky Way, 319; neutron star merger with, 264; primordial, 125; from supernova, 266
blazars, 261–63; diffuse high-energy neutrino flux from, *260*; ultra-high energy cosmic rays and, 249, *249*. *See also* active galactic nuclei (AGN)
BL Lacertae objects, 261
B-mode polarization of CMB, 101–2, 220
Boltzmann equation, 57–63
Boltzmann equation for number density, 58; of baryon number, 192; of beryllium-7, 117; collision terms in, 58–61; in dark matter coannihilations, 136–39; entropy density and, 62–63; of free electrons during recombination, 61–62; of frozen-in dark matter, 141–42; of hydrogen-3 plus helium-3 nuclei, 116; for ionized fraction of hydrogen, 82; matter-antimatter asymmetry and, 139; of neutrons, 111; of stable dark matter candidate, 129–30; of stable Majorana fermion, *132*, 133; of sterile neutrinos, 284; of WIMP, 62
Bose-Einstein distribution, 48
Bose enhancement, 59
bosons of Standard Model, 30; pressure and energy density, 18–19
branching fraction, 43
bremsstrahlung, 65–67; in electromagnetic calorimeter, 329; gamma rays from cosmic

rays and, 255, *255*; synchrotron radiation compared to, 67, 69
bullet cluster, 127

calorimeters of LHC, *329*, 329–30
C and CP violation: in GUTs, 185, 187, 188; in leptogenesis, 195, 196; second Sakharov condition and, 184, 188, 190; by weak interactions, 185. *See also* CP violation
carbon-12 nuclei: in hypothetical equilibrium, 113–14, *114*; stellar nucleosynthesis of, 116
carbon nuclei, spallation to boron, 243–44
ΛCDM cosmological model, 8; angular power spectrum of CMB and, 92, *93*; polarization of CMB and, 102, *103*
"chameleon" scalar fields, 234
charginos, 274, 278–79, 281
chemical equilibrium, 46; chemical potentials in, 51–52; of electrons, positrons, and photons, 52–53; neutron-to-proton ratio and, 110; number density in, 47
chemical potentials, 51–53; dark matter–dark antimatter asymmetry and, 139–40; of decoupled particles, 57; defined, 51; in phase space distribution function, 47; sphaleron transitions and, 191; for Standard Model particles after sphaleron washout, 192–94
Cherenkov telescopes, 316
Christoffel symbols, 15–16
CKM (Cabibbo-Kobayashi-Maskawa) matrix, 34, 36, 176
coincidence problem, 231, 233
cold dark matter (CDM): about 85% of universe's matter density, 125; abundance from CMB measurements, 103; coldness as requirement for, 127; smallest halos forming first, 167; thermal relic freeze out, 131–33, *132*; transfer function for, 159, 166, *166*, 168–69, *169*
Coleman-Mandula theorem, 272n
colliders. *See* Large Hadron Collider (LHC); particle accelerators
color anomaly, 289, 291
color charge, 31, 33–34, 36, 37
color confinement, 37
comoving distance, 26, *27*
comoving horizon, 26; density perturbations and, 165
comoving wavenumber of Fourier mode, 92
Compton scattering: of CMB photons in galaxy clusters, 98; cross section for, 43–44; squared amplitude for, 40
conservation laws, 35–37, 271–72
correlation functions, 158–59
cosmic infrared background: cosmic rays scattering on, 266; pair production in scattering of gamma rays by, 75, 256, *256*; scattering of nuclei with, 242, 250–51; ultra-high energy nuclei scattering on, 266
cosmic microwave background (CMB), 86–107; blackbody spectrum of, 86–88, *87*; cosmic rays scattering on, 266; cosmic strings and, 205–6; current state of measurements, 102–4; curvature of universe and, 27, 98, *99*, 100; dark energy and, 231, 234, 236; dark matter annihilation and, 315–16, *320*, 322, *322*; dark matter decay and, 285; energy density of, 88; energy density variations in, 88, 94; horizon problem and, 211, *212*; inflation and, 226; inverse Compton scattering from, 70; isotropic scattering of photons of, 98; measured properties of, 7; non-Gaussianities in, 226; number density of, 88; optical depth of, 83–84; pair production in scattering of gamma rays by, 75, 256, *256*; photon decoupling and, 6, 54; polarization of, 101–2, 220; primordial density perturbations in, 218–20; recombination and, 79; scattering of free electrons with, 83–84; scattering of gamma rays with, 75; scattering of nuclei with, 250–51; scattering of protons with, 242, 250, 251; surface of last scattering and, 84–85, 86; temperature of, 86, 87. *See also* angular power spectrum of CMB; anisotropies of CMB
cosmic neutrino background, 5, 56
cosmic-ray electrons, 241, 253–54; bremsstrahlung produced by, 67; cooling mechanisms for, 67; distance from origin to energy loss, 244; inverse Compton scattering of, 69–71, *71*, 244;synchrotron emission from, 69, 244
cosmic rays, 241–51; acceleration of, 245–48; anisotropy recently reported in, 251; antiprotons and antinuclei in, *320*, 321, 322–23; background to dark matter searches and, 319, *320*; from blazars, 262–63; from dark matter annihilation, 320–23, *321*, *322*; isotropic arrival of, 241, 249–50; maximum energy of acceleration, 248; measurement of, 241–42; neutrinos produced in interactions of, 179, 258–60, *260*; positrons in, *321*, 321–22, *322*; propagation of, 242–44; protons or nuclei in, 71–74, 241, 248–51, *249*, 259, 266; secondary production of, 321; secondary-to-primary ratios, 244, 323;

cosmic rays (*continued*)
sources of, 241, 242, *249*, 249–50, 251; spectrum of, 241–42, *242*, 246, 247–48; ultra-high energy, 248–51, *249*, 266; in vicinity of galactic disk, 244. *See also* cosmic-ray electrons
cosmic shear, 170, *170*
cosmic strings, 204–6, *205*; axions and, 298–300
cosmic variance, 91, *93*
cosmological constant Λ, 20; anthropic principle and, 237; dark energy and, 9, 21, 24, 102, 230–31, 235–36, *236*; energy density and, 22
cosmological horizon, 26; acoustic oscillations and, 95; density perturbations and, 92; size of, 211; at time of last scattering, 94. *See also* horizon problem
cosmological term, 20
cosmology: ΛCDM model, 8, 92, *93*, 102, *103*; early science of, 3; most important data of, 6–8; open questions and puzzles in, 8–10
CPT theorem, 60n; Boltzmann equation and, 60
CP violation: decay of right-handed neutrinos and, 196; electroweak baryogenesis and, 202; neutrino oscillations and, 178, 179; in QCD, 9–10, 288; second Sakharov condition and, 184, 202. *See also* C and CP violation
critical energy density, 22–23; CMB energy as fraction of, 88; flatness of our universe and, 102
curvature: angular diameter distance and, 26–27; density perturbations and, 157, 159; of Einstein's static universe, 20; Einstein tensor and, 11–12; energy density and, 22–23; evolution of scale factor and, 23–24, *24*; flatness problem and, 210–11; inflation and, 214; luminosity distance and, 27–28; spacetime metric and, 12–14; of universe from CMB measurements, 98, *99*, 100
curvature perturbations, 170–71, 297

damped Lyman-alpha systems, 118
damping tail of CMB, 97–98; dark radiation and, 104, 121
dark energy, 230–38; accelerating expansion and, 231; anthropic principle and, 236–37; beginning of domination of, 6, 25, *25*; cosmological constant and, 9, 21, 24, 102, 230–31, 235–36, *236*; density perturbations and, 156–57; de Sitter universe and, 24; dynamical models of, 231–34, 236, 237; equation of state, 232, 234, 235–36, *236*; evolution of our universe and, 24–25, *25*; ISW rise and, 95; as largest component of energy density, 102; observational probes of, 234–36, *235*; puzzling density of, 9
dark matter, 124–49; about 85% of matter density, 125; abundance from CMB measurements, *99*, 100–101, 103, 323; asymmetric, 139–41; constraints on mass of, 128; decay products of, 323–24; density perturbations in, 155, 156–57; direct searches for, 278, 280, 281, 302–13; dynamical evidence for, 124–25; freeze-in scenario for, 141–43, *144*; in hidden sectors, 300; hot, warm, and cold, 164–68, *166*; indirect searches for, 314–25; IR-dominated, 142; Jeans length and, 154–55, 166–67; large-scale structure and, 7, 323; lifetime of particles in, 323–24; MACHOs and, 125–27; Majorana candidates, 129, 129n, *132*, 132–33; modifications to gravity and, 127; nonthermal production mechanisms for, 141–45, *146*; nuclei and (*see* dark matter scattering with nuclei); as open question, 9; out-of-equilibrium decay scenarios for, 143–45, *146*; requirements of candidate for, 127–28; thermal freeze out of, 314–15; too heavy to produce at LHC, 333; transfer function and, 164–68, *166*; UV-dominated, 142. *See also* cold dark matter (CDM); hot dark matter; thermal relics; warm dark matter; WIMPs (weakly interacting massive particles)
dark matter annihilation, 315–23; to antiprotons and antinuclei, *320*, 321, 322–23; to charged leptons, 322, *322*; constraints on cross section of, 315–16, 319, *320*, 322, *322*, 323; cosmic microwave background and, 315–16, *320*; cosmic rays from, 320–23, *321*, *322*; dwarf galaxies and, 319–20, *320*; Galactic Center and, 317–19, *318*; gamma rays from, 316–20; indirect searches motivated by, 314–15; rate per volume, 316–17
dark matter halos: gravitational collapse and, 159–64; hot dark matter and, 167; matter distribution within, 317–18; of Milky Way, 317–18, *318*, 321, 324; subhalos of Milky Way, 319
dark matter scattering with nuclei, 302–4, *303*; cross sections of, 304–8; direct detection constraints from, 306, 308–11, *309*; neutrino floor for detection of, 308, *309*; spin-dependent interactions, 305, 308, *309*, 311; spin-independent interactions, 305–6, 308, *309*, 309–10
dark radiation, 104–6, *106*, 119, 121

decay rates, 39–40, 42–43
decay width, 43
decoupling, 53–54; evolution after, 56–57, 86; of neutrinos, 5, 54–56, 61. *See also* photon decoupling
degrees of freedom: internal, 47, *47*; relativistic, 49–51, *50*
delta baryon, and pion production, 74, 255, 262–63, 265, 266
density perturbations: acoustic oscillations and, 95–96, 153; adiabatic, 170–71; dark energy and, 156–57; in dark matter, 155, 156–57; evolution of, 92–94, 155–57, 156n; evolution of collapsing region, 159–61, *161*; in first few hundred thousand years, 150; Fourier analysis of, 92–94, 95–96, 158–59; free streaming neutrinos and, 97–98; halo formation in, 161–64; isocurvature, 170–71; statistical properties of, 157–59. *See also* gravitational collapse; primordial density perturbations
de Sitter universe, 24
detailed balance, principle of, 61
deuterium, primordial abundance of, 118, 119, *120*, 121, 122
deuteron formation, 114–16, *115*
DFSZ (Dine-Fisher-Srednicki-Zhitnisky) model, 291, 300
diffusion coefficient of cosmic rays, 242–44, 264
diffusion damping, 97–98; transfer function and, 166
diffusion-loss equations, 243, 321
diffusive shock acceleration, 246–48
Dirac dark matter candidates, 129–30, 129n, 132, 140
Dirac neutrino masses, 174–75, 176–77, 195
direct detection of dark matter, 302–13; lightest neutralino and, 278, 281; pure wino and, 280; ruling out many neutralino models, 281
distance, definitions of, 25–28
Dodelson-Widrow mechanism, 282, 285, 286, *286*, 287
domain walls, 202–4; axions and, 298, 300
dwarf galaxies: baryonic Jeans mass and, 154; dark matter annihilation and, 319–20, *320*; reionization and, 83. *See also* Milky Way, satellite galaxies of
dwarf stars, primordial lithium abundance in, 118

earth's motion, and CMB temperature, 88
effective field theory, 304
effective number of neutrino species, 104–6, *106*, 119, 121
Einasto profile, 317
Einstein-de Sitter universe, 23
Einstein equations, 11–12; with cosmological term, 20; Friedmann equations and, 15, 17
Einstein radius, 125–26
Einstein tensor, 11–12
elastic scattering, Feynman diagrams of, 32–33, 34–35
electric charge, 36, 37, 38
electric fields: acceleration of charged particle and, 248; bremsstrahlung and, 65–66, 253
electromagnetic calorimeter of LHC, *329*, 329–30
electromagnetic force: exchange of photons in, 30; large-scale structure and, 150, 168; range of, 40; screening by virtual particles and, 37
electron-positron colliders, 328–29
electron-positron elastic scattering, 32
electron-positron pair annihilation, 32–33
electron-positron pair production. *See* pair production, electron-positron
electrons: in equilibrium with positrons, 51–53, *53*; number density during recombination, 61–62; scattering of photons by, 79–80; in Standard Model, 31
electrons, high-energy interactions, 65; bremsstrahlung, 65–67; Compton scattering in galaxy clusters, 98; inverse Compton scattering, 69–71, *71*; synchrotron radiation, 67–69. *See also* cosmic-ray electrons
electroweak baryogenesis, 197–202
electroweak gauge bosons, 38. See $W^{\pm}$ bosons; Z boson
electroweak phase transition, 5, 37–39; baryogenesis and, 197, 200–202; departures from equilibrium and, 53–54, 197; not strongly first order in Standard Model, 197; sphalerons and, 201
electroweak symmetry breaking, 38, 173; Higgs boson and, 187; higgsino-like neutralinos and, 279
E-mode polarization of CMB, 101
energy density: in chemical equilibrium, 48–49; of dark radiation, 104–5, 119, 121; in energy-momentum tensor, 17; of ensemble of species, 49; in equilibrium with no chemical potential, 18–19; evolving in our universe, 24–25; expansion of universe and, 21–22; Friedmann equations and, 17–18
energy-momentum four-vector, 41
energy-momentum tensor, 12, 17
entropy, 50–51; in Boltzmann equation, 59–60; relativistic degrees of freedom in, 50, *50*

entropy conservation: for dark matter candidates, 130, 142, 154–55; in equilibrium, 51; misalignment axions and, 296; thermal axions and, 292–93; at time of neutrino decoupling, 55
entropy density, 50–51; axion strings and, 298–99; in Boltzmann equation, 62–63; net baryon density and, 184, 188
entropy dumps, 51, 56, 106; number density of baryons to photons and, 184
equation of state, 19; accelerated expansion and, 213, 231; with constant energy density, 213–14; of dark energy, 232, 234, 235–36, *236*; evolution of energy density and, 22; flatness problem and, 211; heavy relics and, 213; horizon problem and, 212; of ideal gas, 152; of inflaton, 216; isocurvature perturbations and, 171; ISW effect and, 95; during radiation domination era, 153; sound horizon and, 96
equilibrium, 46–50; Boltzmann equation and, 57–58, 60–61; after decoupling, 57; departures from, 53–54; leptogenesis and, 194, 195–96; scattering cross section in early universe and, 129; third Sakharov condition and, 184, 187, 196
equilibrium condition, 53
Euclidean metric, 12
expansion of universe: accelerating, 231; energy density and, 21–22; entropy dump and, 51; evolution of structure and, 155–57; during formation of light nuclei, 114. *See also* Hubble rate; scale factor

Fermi coupling constant, 39
Fermi-Dirac distribution, 48
Fermi Gamma-Ray Space Telescope, 316, 319, 320
fermions: pressure and energy density, 18–19; of Standard Model, 31
Feynman diagrams, 31–35; amplitudes and, 33, 39–40; conservation laws and, 35–36
Feynman rules, 39
fifth forces, 234
filtering mass, 156n
fine structure constant, 39, 66
first Friedmann equation, 17, 22; Big Bang nucleosynthesis and, 119; dark radiation and, 104; deducing very early T∼100 GeV thermal bath, 332; dynamical dark energy and, 233, 234; energy density during inflation and, 298; evolution of collapsing region and, 161; evolution of density perturbations and, 157; inflation and, 214
first order Fermi acceleration, 246–48, 253
first Sakharov condition, 201
fixed target experiments, 326–27
flatness of our universe, 9, 27, 226; baryon acoustic oscillations and, 102; critical energy density and, 22–23, 102; inflation and, 4, 102, 210–11, 214
flat-spectrum radio quasars (FSRQs), 261
flavor, conservation of, 36
Ford, Kent, 124–25
form factor, nuclear, 305–6, 308
Fourier modes: angular power spectrum and, 92–94, 95–96; matter power spectrum and, 158–59
four-momentum, 41
four-vectors, 40–42
free energy, and sphaleron transitions, 191
freeze in of dark matter, 141–43, *144*
Friedmann equations, 15–18, 20. *See also* first Friedmann equation; second Friedmann equation
Friedmann-Lemaître-Robertson-Walker metric, 13–14

galaxies: distribution of, 7; dynamics suggesting dark matter, 125; formation of, 6; hot dark matter and, 167; uniform matter transformed into, 156, 157; virialization of, 163
galaxy clusters: "bullet cluster" collision of, 127; dark energy and, 235; distribution of, 7; hot dark matter and, 167; matter power spectrum and, 170, *170*; Sunyaev-Zel'dovich effect and, 98; ultra-high energy cosmic rays and, 249, *249*; uniform matter transformed into, 156; velocity dispersion in, 124; virialization of, 163
gamma-ray bursts (GRB), 264–65; cosmic-ray acceleration and, 247; ultra-high energy cosmic rays and, 249, *249*
gamma rays: from active galactic nuclei, 261–62; from blazars, 261–62; from bremsstrahlung, 65, 253, 255, *255*; from dark matter annihilation, 316–20; from decaying dark matter, 324; from decay of neutral pions, 72; electron-positron pair production and, 75; from hadronic processes, 254–55, 257; importance for astronomy, 252; from inverse Compton scattering, 69–71, *71*, 75, 253–54, 255, *255*; isotropic cascade of, 255–57; from leptonic processes, 253–54; observed excess from Galactic Center, 319, *320*; optical depth

of, 75; from pulsars, 267–69; from supernova remnants, 267
Gamow, George, 108
gas: bremsstrahlung and, 66–67; interacting with cosmic-ray protons and nuclei, 71
gauge bosons: dark matter annihilations to, 316, 323; massless, 38; number for given gauge group, 186; *X* and *Y* in GUTs, 186, 186n, 187. See also $W^{\pm}$ bosons; *Z* boson
gauge symmetry, 31; as internal symmetry, 271; of Standard Model, 38
gauge symmetry groups: in GUTs, 185–87, 188; in Standard Model, 38, 185
general relativity: supersymmetry and, 272. *See also* Einstein equations; Friedmann equations
geodesics, 15
gluons: color charge and, 33–34; in dark matter scattering with nuclei, 307, *307*; exchanged in strong force, 30; in Feynman diagrams, 33–34; in proton colliders, 328–29; virtual, 328n
Goldstone bosons, 38
Goldstone's theorem, 289
grand unified theories (GUTs), 4; first Sakharov condition satisfied by, 186; gauge coupling unification in, 272; gauge symmetry groups in, 185–87, 188; magnetic monopoles in, 206–7, 212; neutrino masses and, 175; *X* and *Y* gauge bosons in, 186–90, 186n. *See also* GUT baryogenesis
gravitational collapse, 150–52; linear theory of overdense region, 159–61, *161*; pressure pushing against, 152–55
gravitational lensing: of CMB anisotropies, 234n; of CMB photons, 98, 170, *170*; CMB polarization and, 102; of cosmic strings, 206; dark energy and, 235; of distant galaxies by large-scale structure, 170, *170*
gravitational microlensing, 125–27
gravitational waves: CMB polarization and, 102; cosmic strings and, 206; generated during inflation, 102, 220; strongly first order phase transitions and, 197
gravitinos, 274
gravitons, 105
gravity: density perturbations and, 92–94, 101, 150–52; expansion of universe and, 152; integrated Sachs-Wolfe effect and, 95; modifications to, 127, 169
Gunn-Peterson trough, 83
GUT baryogenesis, 185–90; challenges to, 190; electroweak sphalerons and, 192; leptogenesis and, 194–95. *See also* grand unified theories (GUTs)
GZK (Greisen-Zatsepin-Kuzmin) cutoff, 242, 250, 258

hadronic calorimeter of LHC, *329*, 329–30
hadronic sources of gamma rays, 254–55, 257
hadrons, 37
halo formation, 161–64. *See also* dark matter halos
Harrison-Peebles-Zel'dovich spectrum, 94, 159
heavy relics problem, 212–13; inflation and, 214–15
Heisenberg uncertainty principle: decay width and, 43; resolution achieved in collisions and, 326; virtual particles and, 40
helium: formed before hydrogen, 81; primordial mass fraction (Y_p) of, 80, 112, 116, 118, *120*, 121; reionized, 82
helium-3, primordial abundance of, 118, 119, *120*
helium-3 nuclei, 114–16, *115*; in lithium and beryllium production, 117
helium-4 nuclei: formation of, 115, *115*, 116; in hypothetical equilibrium, 113–14, *114*; in lithium and beryllium production, 117–18
Helm form factor, 306
hidden photons, 105
hidden sectors, 333; dark matter in, 300
hierarchy problem, 9, 272–73, *273*
Higgs bosons, 34, 38; cosmic strings and, 204; dark matter annihilations to, 316, 323; DFSZ model and, 291; discovery of, 331; event rates at LHC and, 331–32; extended supersymmetric sector of, 273–74; in GUTs, 187; masses of Standard Model fermions and, 34, 38; monopoles and, 206; puzzling mass of, 9, 272–73, *273*; right-handed neutrinos and, 195; sphaleron washout and, 193
higgsinos, 274–75, 281
Higgs mechanism, 5, 38, 39, 173–74
Higgs portal, 300
Higgs potential, and electroweak phase transition, 200–201
Higgs vacuum expectation value (vev), 38, 39; neutrino masses and, 195, 197
high-energy astrophysics: electron interactions, 65–71, 98; nucleons and nuclei in, 71–74; photon interactions, 74–75
Hillas plot, 249, *249*

homogeneous and isotropic universe: confirmed by observations, 12; energy-momentum tensor of, 17; horizon problem and, 211–12, *212*; matter power spectrum of, 158; metrics for, 12–14; scale factor of, 15
horizon, cosmological. *See* cosmological horizon
horizon problem, 9, 211–12, *212*; inflation and, 214
hot dark matter, 166–68. *See also* dark matter; thermal relics
Hubble constant, 22
Hubble rate, 22; dark energy and, 231, 235; dark matter equilibrium and, 139, 141; equilibrium and, 46, 53–54, 57; formation of light nuclei and, 115, 116, 117–18; in important models, 22–24; ionization fraction of hydrogen and, 82, *83*; Jeans length and, 153; measurements of, 6–7; neutrino decoupling and, 55; neutron-to-proton ratio and, 111–12; number density of particle species and, 58; of our universe, 24–25; in radiation-dominated era, 25, 109; scattering of photons by free electrons and, 80; slowing growth of density perturbations, 156, 156n; slow roll inflation and, 216–18; temperature and, 51
Hubble time, 142
hydrogen-3 nuclei, 114–16, *115*
hydrogen atoms: excited states of, 82, 83; formed during recombination, 61–62, 82; intergalactic, quasar light absorbed by, 83; ionized fraction of, 81–82, *83*; reionized, 82–83
hypercharge Y, 38; after sphaleron washout, 193

IceCube, 260, *260*, 263, 264, 265
inflation, 210–29; axions and, 297–98, 300; diluting monopole abundance, 207; energy density during, 298; energy density of bath following, 225; eternal, 226–27; flatness of our universe and, 4, 102, 210–11, 214; gravitational waves produced during, 102, 220; horizon problem and, 214; large-field models of, 220–23, *221*, *222*, 226; motivations for, 210–13; Peccei-Quinn symmetry breaking and, 297–98; present and future probes of, 226; primordial B-modes and, 102, 220; primordial density perturbations and, 95, 104, 218–20, 222, *222*, 223, 226; reheating subsequent to, 224–26, 298; small-field models of, 220, 223–24, *224*; solution of accelerated expansion, 213–15; tensor-to-scalar ratio, 220. *See also* slow-roll scalar inflaton field
inflationary multiverse, 227
inflaton, 215–18; couplings with Standard Model particles, 225; decay of, 225; energy density of, 220. *See also* slow-roll scalar inflaton field
infrared radiation: inverse Compton scattering from, 70. *See also* cosmic infrared background
instantons, 190, 192
integrated Sachs-Wolfe (ISW) effect, 95, 234n
intergalactic medium, quasar light in, 83
internal degrees of freedom, 47, *47*
interstellar medium: deuterium in, 118; gamma rays from cosmic rays in, 255, *255*; inverse Compton scattering in, 70; secondary production of cosmic rays in, 321; shells from gamma-ray bursts and, 264–65; shock waves in, 247, 267; synchrotron energy losses in, 67
inverse Compton scattering, 69–71, *71*, 75; cycle of pair production and, 256–57; gamma rays from, 69–71, *71*, 75, 253–54, 255, *255*; pulsar gamma-ray emission and, 269; supernova remnants and, 267
ionization history of early universe, 81–82
IR-dominated dark matter, 142
isocurvature perturbations, 170–71, 297
isospin, weak, 38
isotropic scattering of CMB photons, 98
isotropic universe. *See* homogeneous and isotropic universe
ISW rise, 95

Jeans length, 152–55, 156, 156n; dark matter and, 154–55, 166–67; modes larger than, 165
Jeans mass: baryonic, 154; of hot dark matter, 167, 168
jets: of active galactic nuclei, 261–64, *262*; in proton colliders, 329, 329n
J-factors, 317, *318*, 318–19, 320

k-essence models, 233
kinetic equilibrium, 46; elastic scattering and, 53–54
Klein-Nishina cross section, 70–71, *71*
Klein-Nishina regime, 253, 254, 256–57
KSVZ (Kim-Shifman-Vainshtein-Zahkarov) model, 290–91, 300

Large Electron-Positron (LEP) collider, 328
Large Hadron Collider (LHC): collisions of individual partons in, 328–29; constraints on exotic particles in, 332–33; cross sections at, 330–32, *331*; early universe and, 332–33; event rates at, 330–32; Higgs bosons at, 331–32; particle detection at, *329*, 329–30;

Standard Model particles at, *331*, 331–32; supersymmetric searches and, 281, *331*, 332; synchrotron emission of, 328; total energy available in, 327, 328. *See also* particle accelerators
large-scale structure, 150–72; acoustic oscillations and, 153, 155, 168–70, *169*; adiabatic and isocurvature perturbations, 170–71; axions and, 295; dark energy and, 7, 235; dark matter abundance and, 7, 323; evolving in expanding universe, 155–57, 156n; halo formation in, 161–64; hot, warm, and cold dark matter and, 164–68, *166*; matter power spectrum and, 157–59; Press-Schechter formalism, 163–64; pressure and, 152–55, 156, 156n; sterile neutrinos and, 284–85, 287. *See also* density perturbations; gravitational collapse
leptogenesis: baryogenesis via, 194–97; right-handed neutrinos and, 195–97; sphalerons and, 194–95, 197
leptonic sources of gamma rays, 253–54
lepton number: conservation and, 36; neutrino interactions and, 174; R-parity and, 274. *See also* baryon-minus-lepton number ($B - L$); baryon-plus-lepton number ($B + L$)
leptons, 31; generation of masses of, 173–74
LHC. *See* Large Hadron Collider (LHC)
lifetime of a particle, 43
lightest neutralino, 274–81
lithium: primordial abundance of, 118–19, *120*; produced in BBN, 117, *117*
Lorentz invariants, 41
Lorentz transformation, 41
luminosity, and collision rate at LHC, 330
luminosity distance, *27*, 27–28
Lyman-alpha forest, 170, *170*, 285, 286, 287
Lyman-alpha systems, damped, 118

MACHOs (massive astrophysical compact halo objects), 125–27
magnetic fields: axion coupling to photons and, 290, 311–12; cosmic rays in Milky Way and, 321; encountered by cosmic rays, 242, 248–50, *249*; intergalactic, 250; of Milky Way, 241, 249–50, 251; of neutron star, 267–69; of particle accelerators, 327–28, 329; synchrotron radiation and, 67–69
magnetic monopoles, 206–7, 212
Majorana dark matter candidates, 129, 129n, *132*, 132–33
Majorana neutrino masses, 174–75, 195
Mandelstam variable, 135
matter, energy density of: evolution of, 21–22; measured, 103; pressure and, 18–19
matter-antimatter asymmetry, 9, 182–84. *See also* baryon asymmetry
matter-dominated era, 5, 24–25, *25*
matter power spectrum, 157–59, 168–69, *169*
matter-radiation equality, 25, 157, 165, 169
Maxwell-Boltzmann distribution, 48–49
Mercury's perihelion, precession of, 12
mesons, 37
metrics, for homogeneous isotropic universe, 12–14
metric tensor, 12, 14; Christoffel symbols and, 15–16; inverse of, 15; tensor notation for, 14–15
"Mexican hat" potential, 204, *204*, 289
Milgrom, Mordehai, 127
Milky Way: annihilation *J*-factor in, *318*, 318–19; cosmic-ray electrons of, 69; dark matter halo of, 317–18, *318*, 321, 324; dark matter particle velocities in, 303–4; deuterium in, 118; Galactic Center gamma-ray excess, 319, *320*; Galactic Center of, 317–19, *318*; high-energy positrons from sources in, 321–22; highest energy cosmic rays and, 251; MACHOs detected in halo of, 127; magnetic field of, 241, 249–50, 251; satellite galaxies of, 170, 267, 285, 286, 287, 319–20, *320*; subhalos in, 319; supernovae in, 267
Milky Way-like galaxies, virialization of, 163
Minkowski metric, 13, 41
Modified Newtonian Dynamics (MOND), 127, 169
Møller velocity, 59
monopoles, 206–7
MSSM (minimal supersymmetric standard model), 274
multiverse: anthropic principle and, 236–37; inflationary, 227
muon, 31
muon tracker, *329*, 330

Nambu-Goldstone boson, 289
Navarro-Frenk-White (NFW) profile, 317–18, *318*, 319
negative curvature, metric with, 12–13
neutralinos, 274–75; bino-like, 276–77; example of annihilations and coannihilations, *276*, 276–77; higgsino-like, 277–81, 311; lightest, 274–81; as Majorana fermions, 275–76; wino-like, 278–79, 311
neutral pions, 40, 72, 74
neutrino decoupling, 5, 54–56, 61

neutrino masses, 9, 34n, 173–75; angular power spectrum of CMB and, 97–98, 104; CMB data and, 104; dark matter requirements and, 128, 131; leptogenesis and, 195–97; seesaw mechanism and, 186, 195, 197, 282; sterile neutrinos and, 282
neutrino oscillations, 173, 178–80, 257–58, 282
neutrinos, 31; astrophysical production and propagation of, 257–58; from blazar jet, 261, 263–64; from cosmic ray interactions, 179, 258–60, *260*; cosmogenic, 265–66; created in pion production, 74; diffusion damping and, 97–98; effective number of species of, 104–6, *106*, 119, 121; energy density of, 104; in equilibrium in early universe, 54–55; from gamma-ray bursts, 265; heavier two presumably unstable, 37; as hot dark matter, 168; importance for astronomy, 252; in Large Hadron Collider, 330; mixing of, 174–79, 257–58; from supernovae, 267. *See also* right-handed neutrinos; sterile neutrinos
neutron: decay of, 5, 109, 111, *111*, 112, 114, 116; electric dipole moment of, 288, 289
neutron-proton freeze out, 5, 54, 109–12, *111*, 332
neutron stars: gamma-ray bursts and, 264; at heart of pulsar, 267, 269; from supernova, 266
neutron-to-proton ratio, 109–12, *111*
Noether's theorem, 271
non-Gaussianities in CMB, 226
nonrelativistic matter, 19. *See also* matter
nonrelativistic particles: decoupling and, 57; Maxwell-Boltzmann distribution for, 48–49
N-point correlation functions, 158–59
nuclear form factor, 305–6, 308
nuclear species in BBN: hypothetical equilibrium abundances, 112–14, *114*; important parameters of, 112, *113*. *See also* Big Bang nucleosynthesis (BBN)
nuclei: in cosmic rays, 71, 241, 248–51, *249*, 259, 266; high-energy interactions of, 71–74. *See also* dark matter scattering with nuclei
nucleons: high-energy interactions of, 71–74; particles contained in, 328n
nucleosynthesis. *See* Big Bang nucleosynthesis (BBN); stellar nucleosynthesis

one-forms, 14–15
open universe, 23
optical depth: of CMB, 83–84, 98; of gamma rays, 75, 255–56; of microlensing, 126; of universe, 84
oscillation length, 178, 179
Ostriker, Jerry, 125

pair production, electron-positron: by cosmic-ray protons, 250; cycle of inverse Compton scattering and, 256–57; in dark matter annihilation, 316; Feynman diagrams for, 33; in gamma ray attenuation, 75; gamma rays leading to, 255–56; by pulsars, 267, 269; from *Z* boson, 42–43
partial width, 43
particle accelerators: compared to fixed target experiments, 326–27; constraints on exotic particles in, 332; electromagnetic fields of, 327–28; electron-positron colliders, 328–29; first trillionth of a second and, 4–5, 7–8; linear, 328; proton colliders, 328–33; synchrotron emission at, 327–28; testing electroweak sector, 39; total energy available in, 327. *See also* Large Hadron Collider (LHC)
particle horizon, 26
parton distribution functions, 328n
partons, 328
Pauli blocking, 59, 110
Pauli exclusion principle, 128
Peccei, Roberto, 288
Peccei-Quinn mechanism, 10, 288–92, 297–98, 300
Peebles, Jim, 125
Penzias, Arno, 86
phase space distribution function, 47
phase transitions, 197–200, *199*
photo-hadronic processes, 72–74, *74*, 254–55
photon-baryon fluid. *See* baryon-photon fluid
photon decoupling, 6, 54, 79, 80, 84–85; Boltzmann equation and, 61; equilibrium distribution at time of, 86; gravitational collapse of baryons and, 153
photons: created from electroweak gauge bosons, 38; degrees of freedom, 47; electromagnetic force and, 30; from high-energy electrons, 65–71; high-energy interactions of, 74–75; pressure provided by, 153; scattering off charged particles, 79–80; zero chemical potential of, 52
photon temperature: energy density of ensemble of species and, 49; entropy density and, *50*, 50–51
pion, neutral, 40
pion production: in cosmic-ray air showers, 241; cosmic-ray protons and, 41–42, 71–74, 250, 254–55, *255*, 262–64, 266; in dark matter

annihilation, 316; differential cross section, 74; from gamma-ray bursts, 265; from hadronic cosmic rays colliding with gas, 321; high-energy neutrinos from decays of, 257, 259, 260; in photo-hadronic interactions, 72–74, *74*, 254–55; in proton-proton collisions, 254; supernova remnants and, 267; through hadronic resonances, 74, *74*
pitch angle, 67
Planck mass, 11n
Planck scale, 4
PMNS (Pontecorvo-Maki-Nakagawa-Sakata) matrix, 177, 257
portal couplings, 300
position-time four-vector, 40–41
positive curvature, metric with, 12–13
positrons: in cosmic rays, *321*, 321–22, *322*; high-energy interactions of, 65n
Press-Schechter formalism, 163–64
Press-Schechter mass function, 164
pressure, 18–19; in chemical equilibrium, 48–49; in energy-momentum tensor, 17; of fluid before photon decoupling, 93, 95; Friedmann equations and, 17–18, 49; isocurvature perturbations and, 171; large-scale structure and, 152–55, 156, 156n, 157
primordial density perturbations, 94–95, 159, 164, 169–71, *170*; inflation and, 95, 104, 218–20, 222, *222*, 223, 226; nearly scale-invariant, 103–4
propagators, 39–40
proton decay: constraints on lifetime, 186; in GUTs, 186, 187; *R*-parity and, 274; in Standard Model, 36
proton-neutron freeze out, 5, 54, 109–12, *111*, 332
proton-photon scattering: in AGN jet, 261–63; with cosmic microwave background, 266; gamma-ray bursts and, 265; pion production and, 73–74, *74*
proton-proton scattering: gamma rays and neutrinos from, 264; pion-producing, 72, *73*. *See also* Large Hadron Collider (LHC)
protons: bremsstrahlung and, 66; particles contained in, 328n; scattering with CMB, 242, 250
pulsar braking index, 268
pulsars, 267–69; constraints on cosmic strings and, 206; gamma rays from, 319; from revived neutron star, 269; spindown timescale of, 269; time evolution of, 268–69; ultra-high energy cosmic rays and, 249, *249*; very high-energy positrons accelerated by, 322
p-wave suppressed dark matter candidates, 134

QCD (quantum chromodynamics): complexities of proton-proton collisions and, 328–29; CP violation in, 9–10, 288–89; Feynman diagrams for, 33–34. *See also* strong force
QCD axions, 291–92, 312. *See also* axions
QCD phase transition, 5, 37; relativistic degrees of freedom and, 50
QED (quantum electrodynamics): Feynman diagrams for, 31–33, 34–35
quantum field theory: anomalies in, 289; gauge symmetry in, 31, 271; particle as excitation in, 30
quantum gravity, 4, 10, 272
quantum tunneling: in first order phase transitions, 197, 200, 201; vacuum transitions and, 190
quark-gluon plasma, 4
quarks, 31; color charge of, 33; dark matter annihilations to, 316, 323; in dark matter scattering with nuclei, 304, 307, *307*, 308–11; degrees of freedom, 47; generation of masses of, 173–74; in proton colliders, 328–29; valence quarks vs. sea quarks, 307, 328n
quasars: as active galactic nuclei, 261; gravitational lensing by, 126; reionization by, 83; reionized hydrogen measurement and, 83
Quinn, Helen, 288. *See also* Peccei-Quinn mechanism
quintessence, 231–34, 237

radiation: evolution of energy density of, 21–22; pressure and energy density of, 18–19; relativistic particles as, 19
radiation-dominated era, 24–25, *25*; Hubble rate during, 25, 109
radio emission: axion detection and, 311–12; from pulsar, 267, 269; from synchrotron mechanism in Milky Way, 69
recombination: dark matter annihilation during, 315–16; number density of free electrons during, 61–62. *See also* atoms
redshift, cosmological: definitions of distance and, 26–28, *27*; energy density of radiation and, 22; energy of a relativistic particle and, 56–57; wavelength and, 22
reduced Planck units, 216n
reionization, 6, 82–84
relativistic degrees of freedom, 49–51, *50*

relativistic particles: decoupling and, 56–57; in equilibrium, 48; as "radiation," 19. *See also* radiation
renormalization, 31
Ricci scalar, 12, 17
Ricci tensor, 12, 16–17
right-handed neutrinos, 195–97; hidden sector state mixing with, 300; leptogenesis and, 195–97, *196*; in *SO*(10), 186
rigidity of a particle, 244
R-parity, 274
Rubin, Vera, 124–25

Sachs-Wolfe effect, 88, 94; integrated (ISW), 95, 234n
Sachs-Wolfe plateau, 95
Saha equation, 81–82, *83*
Sakharov conditions, 184–85, 186, 187. *See also* first Sakharov condition; second Sakharov condition; third Sakharov condition
scale factor, 13; Boltzmann equation and, 62; of collapsing overdense region, 159–60; decoupled particles and, 57; density perturbations and, 156–57; dynamical dark energy and, 233; of Einstein-de Sitter universe, 23; entropy density and, 51; evolving in matter-dominated universe, 23–24, *24*; evolving in our universe, 25, *25*; Friedmann equations and, 15–18; luminosity distance and, 27
scattering cross sections, 39–40, 43–44
s-channel diagrams, 33
s-channel exchange, of spin-1 dark matter mediator, 142
s-channel resonance, dark matter annihilation through, 135–36
sea quarks, 307, 328n
second Friedmann equation, 17–18; accelerated expansion and, 213; slow roll dark energy and, 232; slow roll inflation and, 216
second order Fermi acceleration, 245–46
second order phase transition, 197; transitioning out of inflation, 215
second Sakharov condition, 184; electroweak baryogenesis and, 202
seesaw mechanism, 186, 195, 197, 282
Shi-Fuller mechanism, *286*, 287
shock waves: cosmic ray acceleration and, 246–48, 264–65; from gamma-ray bursts, 264–65; from supernova remnants, 267
Silk damping, 97–98, 121; transfer function and, 166
slow-roll parameters, 216–18, 220
slow-roll scalar dynamical dark energy, 231–34
slow-roll scalar inflaton field, 215–18; large-field model, 220–21, *221*; primordial density perturbations and, 218–20, 222, *222*; quantum fluctuations in, 227
sneutrinos, 271, 274
SO(10), 186, 188; electroweak sphalerons and, 192
Sommerfeld enhancements, 279–80
sound horizon, 96, 98, 168–69, 235
sound waves. *See* acoustic oscillations
sphalerons, electroweak, 36; baryon asymmetry and, 185, 190–94, 197, 201, 202; baryon number and, 185; as high-temperature vacuum transitions, 190; leptogenesis and, 194–95, 197
spin states, number of, 47
spontaneous symmetry breaking, 198; of electroweak sector, 38; forming cosmic strings, 204–5, *205*, 298; forming domain walls, 202–4, 298; forming monopoles, 206–7; of $U(1)_{PQ}$ symmetry, 289, 291, 297, 298, 300
Standard Model of particle physics, 30–31; bath of particles a second after Big Bang, 332; gauge symmetry groups of, 38, 185; hidden sector states decaying into, 300; neutrino masses and, 174; only stable particles in, 128; particles produced at LHC, *331*, 331–32
starburst winds, 249, *249*
starlight, inverse Compton scattering from, 70
stars: constraint on axions and, 290; deuterium processed in, 118; formation of, 6; helium-3 in, 118; helium produced in, 118; lithium deletion in, 119; population III, and reionization, 83
static universes, 19–21
staus, in neutralino annihilations and coannihilations, *276*, 276–77
steady state models, 108
stellar nucleosynthesis, 108–9, 116
sterile neutrinos, 105, 174–75, 180, 281–88; decays of, 285–86, *286*; lepton asymmetry and, *286*, 286–88; mass matrix including active neutrino and, 282; number density of, 284, *285*; produced through oscillations, 282–86, *286*, 287; X-ray signal from decays of, 286, *286*
stops, light, 201
strings, cosmic, 204–6, *205*; axions and, 298–300
string theory, 205n, 272
strong CP problem, 9–10, 288–89, 291

strong force: binding nucleons into nuclei, 109; exchange of gluons in, 30; range of, 37, 40. *See also* QCD (quantum chromodynamics)
SU(2), 38, 185; sphalerons and, 192
SU(3), 38, 185
SU(5), 185–87, 188; sphalerons and, 192, 194
subhalos in Milky Way, 319
sun: deflection of light by, 12; neutrino oscillations from fusion in, 179; neutrinos propagating through, 180
Sunyaev-Zel'dovich effect, 98, 234n
Super-Kamiokande, 186
supernova 1987A: constraint on axions and, 290, *294*; neutrinos produced by, 257, 267, 290
supernovae, 266–67; Type Ia, and dark energy, 231, 235, *235*, 236
supernova remnants: cosmic-ray acceleration and, 247, 248, 249, *249*, 267; cosmic rays from, 267; gamma rays and neutrinos from, 267, 319; shock waves associated with, 247
supersymmetry, 271–72; both types of symmetry in, 271–72; dark matter and, 272, 273–74; Higgs mass and, 9, 272–73, *273*; LHC searches for, 281, *331*, 332; motivations for studying, 272; potentially viable parameter space, 281; proton decay and, 186
surface of last scattering, 84–85, 86; CMB polarization and, 101–2; horizon problem and, *212*
symmetry, two categories of, 271–72
synchrotron radiation, 67–69; critical frequency of, 67–68; at particle accelerator, 327–28; suppressing gamma-ray emission, 254

taus, 31; in neutralino annihilations and coannihilations, *276*, 276–77
t-channel diagrams, 33, 35
temperature, and distribution of momenta, 46–47
tensor notation, 14–15
tensor-to-scalar ratio, 220, *222*, 222–24, 226
tensor-vector-scalar gravity (TeVeS), 127
TeV halos, 269
thermal relics, 129–39; annihilation cross sections and, 314–15, 316, *322*, 323; annihilations near a resonance, 135–36; axions as, 292–93, *294*; coannihilations of, 136–39; cold relic freeze out, 131–33, *132*; evolution of number density, 129–30; hot relic freeze out, 130–31; hot relics' suppression of structures, 167; Lyman-alpha forest and, 170; neutralinos and, 277, 278, 281; pure wino as, 279–80, 281; sterile neutrinos and, 284–85; velocity-dependent annihilations of, 133–35
third Sakharov condition: bubble nucleation and, 200; electroweak baryogenesis and, 201; equilibrium and, 184, 187, 196; strongly first order phase transitions and, 197, 200
Thomson cross section: for bremsstrahlung, 66; for inverse Compton scattering, 70, 71, 253; for photon scattering by electrons, 80; for synchrotron radiation, 67
Thomson limit, 70, 244
Thomson regime, gamma rays produced in, 253–54
Thomson scattering: diffusion and, 97; polarization of CMB and, 101–2. *See also* Thomson cross section
time of last scattering, 88, 94, 96–98, 100
topological defects, 202–7; axion production from, *294*, 298–300; strongly first order phase transitions and, 197
tracker of LHC, 329, *329*
transfer function of matter power spectrum, 159; of adiabatic and isocurvature perturbations, 171; baryons' impact on, 168–69, *169*; dark matter and, 164–68, *166*
tree-level Feynman diagrams, 34–35
Tremaine-Gunn bound, 128, 284
triple alpha process, 116
tunneling. *See* quantum tunneling
2-point correlation function, 91–92, 158–59; matter power spectrum and, 169

$U(1)$ symmetry, 38, 185; cosmic strings and, 204, *205*; monopoles and, 206
$U(1)_{PQ}$ symmetry, 288–89
u-channel diagrams, 33, 35
universe, brief history of, 4–6
UV-dominated dark matter, 142

vacuum energy density: anthropic principle and, 236–37; cosmological constant and, 9, 22, 230–31
vacuum transitions, 190–91, *191*. *See also* phase transitions
vectors, and metric tensor, 14
virialization of overdense region, 161–63
virtual particles, 37, 40

warm dark matter, 167–68, 170, 284
wavenumber of Fourier mode, 92, 95–96
Waxman-Bahcall bound, 260, *260*, 263, 265, 266

$W^{\pm}$ bosons, 30; changing type of a fermion, 34; coupling to quarks, 34, 36; masses of, 38, 39, 40
weak force: acting on quarks and leptons, 31; angular momentum and, 31; in conversion between protons and neutrons, 109–12, *111*; exchange of $W^{\pm}$ and Z bosons, 30; Feynman diagrams for, 34, 40; sterile neutrinos and, 174
weak isospin, 38
weak mixing angle, 38
Weinberg, Steven, 237, 289
Weinberg angle, 38
Wheeler, John, 12
Wilczek, Frank, 289
Wilson, Robert, 86
WIMPs (weakly interacting massive particles), 125; evolution and thermal freeze out, 62; gravitational collapse of, 155; scattering with nuclei, 302, 311, 314; stringent constraints on, 300
"wine bottle" potential, 204, *204*, 289
winos, 274–75, 279–80

X and *Y* gauge bosons in GUTs, 186–90, 186n
X decays, 188–89, *189*
X-ray line searches, 286, *286*, 287–88
X-rays from bremsstrahlung, 65

Y decays, 189
Yukawa coupling, 34, 39, 173–74; electroweak phase transition and, 201; in GUTs, 187; neutrino mass and, 195, 196

Z bosons, 30, 34, 38, 40; pair production from, 42–43
Zwicky, Fritz, 124